PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 7 : C12N 15/12, C07K 14/705, C12Q 1/68, A61K 38/17, C07K 16/28, 14/72

A2

(11) International Publication Number: WO 00/15793

(43) International Publication Date: 23 March 2000 (23.03.00)

(21) International Application Number: PCT/US99/20958

(22) International Filing Date: 17 September 1999 (17.09.99)

(30) Priority Data:
09/156,513 17 September 1998 (17.09.98) US

(63) Related by Continuation (CON) or Continuation-in-Part (CIP) to Earlier Application
US 09/156,513 (CIP)
Filed on 17 September 1998 (17.09.98)

(71) Applicant (*for all designated States except US*): INCYTE PHARMACEUTICALS, INC. [US/US]; 3174 Porter Drive, Palo Alto, CA 94304 (US).

(72) Inventors; and
(75) Inventors/Applicants (*for US only*): BANDMAN, Olga [US/US]; 366 Anna Avenue, Mountain View, CA 94043 (US). LAL, Preeti [IN/US]; 2382 Lass Drive, Santa Clara, CA 95054 (US). TANG, Y., Tom [CN/US]; 4230 Ranwick Court, San Jose, CA 95118 (US). CORLEY, Neil, C. [US/US]; 1240 Dale Avenue #30, Mountain View, CA 94040 (US). GUEGLER, Karl, J. [CH/US]; 1048 Oakland Avenue, Menlo Park, CA 94025 (US). GORGONE, Gina, A. [US/US]; 1253 Pinecrest Drive, San Francisco, CA 94132 (US). BAUGHN, Mariah, R. [US/US]; 14244 Santiago Road, San Leandro, CA 94577 (US).

(74) Agents: BILLINGS, Lucy, J. et al.; Incyte Pharmaceuticals, Inc., 3174 Porter Drive, Palo Alto, CA 94304 (US).

(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published
Without international search report and to be republished upon receipt of that report.

(54) Title: HUMAN GPCR PROTEINS

(57) Abstract

The invention provides human GPCR proteins (HGPRP) and polynucleotides which identify and encode HGPRP. The invention also provides expression vectors, host cells, antibodies, agonists, and antagonists. The invention also provides methods for diagnosing, treating or preventing disorders associated with expression of HGPRP.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

HUMAN GPCR PROTEINS

TECHNICAL FIELD

This invention relates to nucleic acid and amino acid sequences of human GPCR proteins and to the use of these sequences in the diagnosis, treatment, and prevention of cell proliferative, neurological, and immune disorders.

BACKGROUND OF THE INVENTION

The term receptor describes proteins that specifically recognize other molecules. The category is broad and includes proteins with a variety of functions. The bulk of the proteins termed receptors are cell surface proteins which bind extracellular ligands, leading to cellular responses including growth, differentiation, endocytosis, and immune response. Other proteins termed receptors facilitate the specific transport of proteins across the endoplasmic reticulum membrane and localize enzymes to a particular location in the cell.

G protein coupled receptors (GPCR) are a superfamily of integral membrane proteins which transduce extracellular signals. GPCRs include receptors for biogenic amines; for lipid mediators of inflammation, peptide hormones, and sensory signal mediators. The GPCR becomes activated when the receptor binds its extracellular ligand. Conformational changes in the GPCR, which result from the ligand-receptor interaction, affect the binding affinity of a G protein to the GPCR intracellular domains. This enables GTP to bind with enhanced affinity to the G protein. Activation of the G protein by GTP leads to the interaction of the G protein α subunit with adenylate cyclase or other second messenger molecule generators. This interaction regulates the activity of adenylate cyclase and hence production of a second messenger molecule, cAMP. cAMP regulates phosphorylation and activation of other intracellular proteins. Alternatively, cellular levels of other second messenger molecules, such as cGMP or eicosinoids, may be upregulated or downregulated by the activity of GPCRs. The G protein α subunit is deactivated by hydrolysis of the GTP by GTPase, and the β, γ, and α subunits reassociate. The heterotrimeric G protein then dissociates from the adenylate cyclase or other second messenger molecule generator. Activity of GPCR may also be regulated by phosphorylation of the intra- and extracellular domains or loops.

Visual excitation and the phototransmission of light signals is a signaling cascade in which GPCRs play an important role. The process begins in retinal rod cells with the absorption of light by the photoreceptor rhodopsin, a GPCR composed of a 40-kDa protein, opsin, and a chromophore, 11-cis-retinal. The photoisomerization of the retinal chromophore causes a

conformational change in the opsin GPCR and activation of the associated G-protein, transducin. This activation leads to the hydrolysis of cyclic-GMP and the closure of cyclic-GMP regulated, Ca^{2+}-specific channels in the plasma membrane of the rod cell. The resultant membrane hyperpolarization generates a nerve signal. Recovery of the dark state of the rod cell involves the activation of guanylate cyclase leading to increased cyclic-GMP levels and the reopening of the Ca^{2+}-specific channels (L. Stryer (1991) J. Biol. Chem. 266:10711-10714).

Glutamate receptors form a group of GPCRs that are important in neurotransmission. Glutamate is the major neurotransmitter in the CNS and is believed to have important roles in neuronal plasticity, cognition, memory, learning and some neurological disorders such as epilepsy, stroke, and neurodegeneration (Watson, S. and S. Arkinstall (1994) The G-Protein Linked Receptor Facts Book, Academic Press, San Diego CA, pp. 130-132).. These effects of glutamate are mediated by two distinct classes of receptors termed ionotropic and metabotropic. Ionotropic receptors contain an intrinsic cation channel and mediate fast, excitatory actions of glutamate. Metabotropic receptors are modulatory, increasing the membrane excitability of neurons by inhibiting calcium dependent potassium conductances and both inhibiting and potentiating excitatory transmission of ionotropic receptors. Metabotropic receptors are classified into five subtypes based on agonist pharmacology and signal transduction pathways and are widely distributed in brain tissues.

The vasoactive intestinal polypeptide (VIP) family is a group of related polypeptides whose actions are also mediated by GPCRs. Key members of this family are VIP itself, secretin, and growth hormone releasing factor (GRF). VIP has a wide profile of physiological actions including relaxation of smooth muscles, stimulation or inhibition of secretion in various tissues, modulation of various immune cell activities. and various excitatory and inhibitory activities in the CNS. Secretin stimulates secretion of enzymes and ions in the pancreas and intestine and is also present in small amounts in the brain. GRF is an important neuroendocrine agent regulating synthesis and release of growth hormone from the anterior pituitary (Watson, S. and S. Arkinstall supra, pp. 278-283).

The structure of GPCRs is highly-conserved and consists of seven hydrophobic transmembrane (serpentine) regions, cysteine disulfide bridges between the second and third extracellular loops, an extracellular N-terminus, and a cytoplasmic C-terminus. Three extracellular loops alternate with three intracellular loops to link the seven transmembrane regions. The most conserved parts of these proteins are the transmembrane regions and the first two cytoplasmic loops. A conserved, acidic-Arg-aromatic residue triplet present in the second cytoplasmic loop may interact with the G-proteins. The consensus pattern of the G-protein

coupled receptors signature (PS00237; SWISSPROT) is characteristic of most proteins belonging to this superfamily (Watson, S. and S. Arkinstall supra, pp. 2-6).

The discovery of new human GPCR proteins and the polynucleotides encoding them satisfies a need in the art by providing new compositions which are useful in the diagnosis, prevention, and treatment of cell proliferative, neurological, and immune disorders.

SUMMARY OF THE INVENTION

The invention features substantially purified polypeptides, human GPCR proteins, referred to collectively as "HGPRP". In one aspect, the invention provides a substantially purified polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-6, and fragments thereof.

The invention further provides a substantially purified variant having at least 90% amino acid identity to at least one of the amino acid sequences selected from the group consisting of SEQ ID NO:1-6, and fragments thereof. The invention also provides an isolated and purified polynucleotide encoding the polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-6, and fragments thereof. The invention also includes an isolated and purified polynucleotide variant having at least 70% polynucleotide sequence identity to the polynucleotide encoding the polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-6, and fragments thereof.

Additionally, the invention provides an isolated and purified polynucleotide which hybridizes under stringent conditions to the polynucleotide encoding the polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-6, and fragments thereof. The invention also provides an isolated and purified polynucleotide having a sequence which is complementary to the polynucleotide encoding the polypeptide comprising the amino acid sequence selected from the group consisting of SEQ ID NO:1-6, and fragments thereof.

The invention also provides an isolated and purified polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:7-12, and fragments thereof. The invention further provides an isolated and purified polynucleotide variant having at least 70% polynucleotide sequence identity to the polynucleotide sequence selected from the group consisting of SEQ ID NO:7-12, and fragments thereof. The invention also provides an isolated and purified polynucleotide having a sequence which is complementary to the polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:7-12, and fragments thereof.

The invention also provides a method for detecting a polynucleotide in a sample containing nucleic acids, the method comprising the steps of: (a) hybridizing the complement of

the polynucleotide sequence to at least one of the polynucleotides of the sample, thereby forming a hybridization complex; and (b) detecting the hybridization complex, wherein the presence of the hybridization complex correlates with the presence of a polynucleotide in the sample. In one aspect, the method further comprises amplifying the polynucleotide prior to hybridization.

The invention further provides an expression vector containing at least a fragment of the polynucleotide encoding the polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-6, and fragments thereof. In another aspect, the expression vector is contained within a host cell.

The invention also provides a method for producing a polypeptide, the method comprising the steps of: (a) culturing the host cell containing an expression vector containing at least a fragment of a polynucleotide under conditions suitable for the expression of the polypeptide; and (b) recovering the polypeptide from the host cell culture.

The invention also provides a pharmaceutical composition comprising a substantially purified polypeptide having the amino acid sequence selected from the group consisting of SEQ ID NO:1-6, and fragments thereof, in conjunction with a suitable pharmaceutical carrier.

The invention further includes a purified antibody which binds to a polypeptide selected from the group consisting of SEQ ID NO:1-6, and fragments thereof. The invention also provides a purified agonist and a purified antagonist to the polypeptide.

The invention also provides a method for treating or preventing a disorder associated with decreased expression or activity of HGPRP, the method comprising administering to a subject in need of such treatment an effective amount of a pharmaceutical composition comprising a substantially purified polypeptide having the amino acid sequence selected from the group consisting of SEQ ID NO:1-6, and fragments thereof, in conjunction with a suitable pharmaceutical carrier.

The invention also provides a method for treating or preventing a disorder associated with increased expression or activity of HGPRP, the method comprising administering to a subject in need of such treatment an effective amount of an antagonist of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-6, and fragments thereof.

BRIEF DESCRIPTION OF THE TABLES

Table 1 shows nucleotide and polypeptide sequence identification numbers (SEQ ID NO), clone identification numbers (clone ID), cDNA libraries, and cDNA fragments used to assemble full-length sequences encoding HGPRP.

Table 2 shows features of each polypeptide sequence including potential motifs, homologous sequences, and methods and algorithms used for identification of HGPRP.

Table 3 shows the tissue-specific expression patterns of each nucleic acid sequence as determined by northern analysis, conditions, diseases or disorders associated with these tissues, and the vector into which each cDNA was cloned.

Table 4 describes the tissues used to construct the cDNA libraries from which Incyte clones encoding HGPRP were isolated.

Table 5 shows the programs, their descriptions, references, and threshold parameters used to analyze HGPRP.

DESCRIPTION OF THE INVENTION

Before the present proteins, nucleotide sequences, and methods are described, it is understood that this invention is not limited to the particular machines, materials and methods described, as these may vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to limit the scope of the present invention which will be limited only by the appended claims.

It must be noted that, as used herein and in the appended claims, the singular forms "a," "an," and "the" include plural reference unless the context clearly dictates otherwise. Thus, for example, a reference to "a host cell" includes a plurality of such host cells, and a reference to "an antibody" is a reference to one or more antibodies and equivalents thereof known to those skilled in the art, and so forth.

Unless defined otherwise, all technical and scientific terms used herein have the same meanings as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any machines, materials, and methods similar or equivalent to those described herein can be used to practice or test the present invention, the preferred machines, materials and methods are now described. All publications mentioned herein are cited for the purpose of describing and disclosing the cell lines, protocols, reagents and vectors which are reported in the publications and which might be used in connection with the invention. Nothing herein is to be construed as an admission that the invention is not entitled to antedate such disclosure by virtue of prior invention.

DEFINITIONS

"HGPRP" refers to the amino acid sequences of substantially purified HGPRP obtained from any species, particularly a mammalian species, including bovine, ovine, porcine, murine, equine, and preferably the human species, from any source, whether natural, synthetic, semi-synthetic, or recombinant.

The term "agonist" refers to a molecule which, when bound to HGPRP, increases or

prolongs the duration of the effect of HGPRP. Agonists may include proteins, nucleic acids, carbohydrates, or any other molecules which bind to and modulate the effect of HGPRP.

An "allelic variant" is an alternative form of the gene encoding HGPRP. Allelic variants may result from at least one mutation in the nucleic acid sequence and may result in altered mRNAs or in polypeptides whose structure or function may or may not be altered. Any given natural or recombinant gene may have none, one, or many allelic forms. Common mutational changes which give rise to allelic variants are generally ascribed to natural deletions, additions, or substitutions of nucleotides. Each of these types of changes may occur alone, or in combination with the others, one or more times in a given sequence.

"Altered" nucleic acid sequences encoding HGPRP include those sequences with deletions, insertions, or substitutions of different nucleotides, resulting in a polynucleotide the same as HGPRP or a polypeptide with at least one functional characteristic of HGPRP. Included within this definition are polymorphisms which may or may not be readily detectable using a particular oligonucleotide probe of the polynucleotide encoding HGPRP, and improper or unexpected hybridization to allelic variants, with a locus other than the normal chromosomal locus for the polynucleotide sequence encoding HGPRP. The encoded protein may also be "altered," and may contain deletions, insertions, or substitutions of amino acid residues which produce a silent change and result in a functionally equivalent HGPRP. Deliberate amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues, as long as the biological or immunological activity of HGPRP is retained. For example, negatively charged amino acids may include aspartic acid and glutamic acid, positively charged amino acids may include lysine and arginine, and amino acids with uncharged polar head groups having similar hydrophilicity values may include leucine, isoleucine, and valine; glycine and alanine; asparagine and glutamine; serine and threonine; and phenylalanine and tyrosine.

The terms "amino acid" or "amino acid sequence" refer to an oligopeptide, peptide, polypeptide, or protein sequence, or a fragment of any of these, and to naturally occurring or synthetic molecules. In this context, "fragments," "immunogenic fragments," or "antigenic fragments" refer to fragments of HGPRP which are preferably at least 5 to about 15 amino acids in length, most preferably at least 14 amino acids, and which retain some biological activity or immunological activity of HGPRP. Where "amino acid sequence" is recited to refer to an amino acid sequence of a naturally occurring protein molecule, "amino acid sequence" and like terms are not meant to limit the amino acid sequence to the complete native amino acid sequence associated with the recited protein molecule.

"Amplification" relates to the production of additional copies of a nucleic acid sequence. Amplification is generally carried out using polymerase chain reaction (PCR) technologies well known in the art.

The term "antagonist" refers to a molecule which, when bound to HGPRP, decreases the amount or the duration of the effect of the biological or immunological activity of HGPRP. Antagonists may include proteins, nucleic acids, carbohydrates, antibodies, or any other molecules which decrease the effect of HGPRP.

The term "antibody" refers to intact molecules as well as to fragments thereof, such as Fab, $F(ab')_2$, and Fv fragments, which are capable of binding the epitopic determinant. Antibodies that bind HGPRP polypeptides can be prepared using intact polypeptides or using fragments containing small peptides of interest as the immunizing antigen. The polypeptide or oligopeptide used to immunize an animal (e.g., a mouse, a rat, or a rabbit) can be derived from the translation of RNA, or synthesized chemically, and can be conjugated to a carrier protein if desired. Commonly used carriers that are chemically coupled to peptides include bovine serum albumin, thyroglobulin, and keyhole limpet hemocyanin (KLH). The coupled peptide is then used to immunize the animal.

The term "antigenic determinant" refers to that fragment of a molecule (i.e., an epitope) that makes contact with a particular antibody. When a protein or a fragment of a protein is used to immunize a host animal, numerous regions of the protein may induce the production of antibodies which bind specifically to antigenic determinants (given regions or three-dimensional structures on the protein). An antigenic determinant may compete with the intact antigen (i.e., the immunogen used to elicit the immune response) for binding to an antibody.

The term "antisense" refers to any composition containing a nucleic acid sequence which is complementary to the "sense" strand of a specific nucleic acid sequence. Antisense molecules may be produced by any method including synthesis or transcription. Once introduced into a cell, the complementary nucleotides combine with natural sequences produced by the cell to form duplexes and to block either transcription or translation. The designation "negative" can refer to the antisense strand, and the designation "positive" can refer to the sense strand.

The term "biologically active" refers to a protein having structural, regulatory, or biochemical functions of a naturally occurring molecule. Likewise, "immunologically active" refers to the capability of the natural, recombinant, or synthetic HGPRP, or of any oligopeptide thereof, to induce a specific immune response in appropriate animals or cells and to bind with specific antibodies.

The terms "complementary" or "complementarity" refer to the natural binding of polynucleotides by base pairing. For example, the sequence "5' A-G-T 3'" bonds to the

complementary sequence "3' T-C-A 5'." Complementarity between two single-stranded molecules may be "partial," such that only some of the nucleic acids bind, or it may be "complete," such that total complementarity exists between the single stranded molecules. The degree of complementarity between nucleic acid strands has significant effects on the efficiency and strength of the hybridization between the nucleic acid strands. This is of particular importance in amplification reactions, which depend upon binding between nucleic acids strands, and in the design and use of peptide nucleic acid (PNA) molecules.

A "composition comprising a given polynucleotide sequence" or a "composition comprising a given amino acid sequence" refer broadly to any composition containing the given polynucleotide or amino acid sequence. The composition may comprise a dry formulation or an aqueous solution. Compositions comprising polynucleotide sequences encoding HGPRP or fragments of HGPRP may be employed as hybridization probes. The probes may be stored in freeze-dried form and may be associated with a stabilizing agent such as a carbohydrate. In hybridizations, the probe may be deployed in an aqueous solution containing salts (e.g., NaCl), detergents (e.g., sodium dodecyl sulfate; SDS), and other components (e.g., Denhardt's solution, dry milk, salmon sperm DNA, etc.).

"Consensus sequence" refers to a nucleic acid sequence which has been resequenced to resolve uncalled bases, extended using the XL-PCR kit (PE Biosystems, Foster City CA) in the 5' and/or the 3' direction, and resequenced, or which has been assembled from the overlapping sequences of more than one Incyte Clone using a computer program for fragment assembly, such as the GELVIEW Fragment Assembly system (GCG, Madison WI). Some sequences have been both extended and assembled to produce the consensus sequence.

The term "correlates with expression of a polynucleotide" indicates that the detection of the presence of nucleic acids, the same or related to a nucleic acid sequence encoding HGPRP, by northern analysis is indicative of the presence of nucleic acids encoding HGPRP in a sample, and thereby correlates with expression of the transcript from the polynucleotide encoding HGPRP.

A "deletion" refers to a change in the amino acid or nucleotide sequence that results in the absence of one or more amino acid residues or nucleotides.

The term "derivative" refers to the chemical modification of a polypeptide sequence, or a polynucleotide sequence. Chemical modifications of a polynucleotide sequence can include, for example, replacement of hydrogen by an alkyl, acyl, or amino group. A derivative polynucleotide encodes a polypeptide which retains at least one biological or immunological function of the natural molecule. A derivative polypeptide is one modified by glycosylation, pegylation, or any similar process that retains at least one biological or immunological function of the polypeptide

from which it was derived.

The term "similarity" refers to a degree of complementarity. There may be partial similarity or complete similarity. The word "identity" may substitute for the word "similarity". A partially complementary sequence that at least partially inhibits an identical sequence from hybridizing to a target nucleic acid is referred to as "substantially similar." The inhibition of hybridization of the completely complementary sequence to the target sequence may be examined using a hybridization assay (Southern or northern blot, solution hybridization, and the like) under conditions of reduced stringency. A substantially similar sequence or hybridization probe will compete for and inhibit the binding of a completely similar (identical) sequence to the target sequence under conditions of reduced stringency. This is not to say that conditions of reduced stringency are such that non-specific binding is permitted, as reduced stringency conditions require that the binding of two sequences to one another be a specific (i.e., a selective) interaction. The absence of non-specific binding may be tested by the use of a second target sequence which lacks even a partial degree of complementarity (e.g., less than about 30% similarity or identity). In the absence of non-specific binding, the substantially similar sequence or probe will not hybridize to the second non-complementary target sequence.

The phrases "percent identity" or "% identity" refer to the percentage of sequence similarity found in a comparison of two or more amino acid or nucleic acid sequences. Percent identity can be determined electronically, e.g., by using the MEGALIGN program (DNASTAR, Madison WI) which creates alignments between two or more sequences according to methods selected by the user, e.g., the clustal method. (See, e.g., Higgins, D.G. and P.M. Sharp (1988) Gene 73:237-244.) The clustal algorithm groups sequences into clusters by examining the distances between all pairs. The clusters are aligned pairwise and then in groups. The percentage similarity between two amino acid sequences, e.g., sequence A and sequence B, is calculated by dividing the length of sequence A, minus the number of gap residues in sequence A, minus the number of gap residues in sequence B, into the sum of the residue matches between sequence A and sequence B, times one hundred. Gaps of low or of no similarity between the two amino acid sequences are not included in determining percentage similarity. Percent identity between nucleic acid sequences can also be counted or calculated by other methods known in the art, e.g., the Jotun Hein method. (See, e.g., Hein, J. (1990) Methods Enzymol. 183:626-645.) Identity between sequences can also be determined by other methods known in the art, e.g., by varying hybridization conditions.

"Human artificial chromosomes" (HACs) are linear microchromosomes which may contain DNA sequences of about 6 kb to 10 Mb in size, and which contain all of the elements

required for stable mitotic chromosome segregation and maintenance.

The term "humanized antibody" refers to antibody molecules in which the amino acid sequence in the non-antigen binding regions has been altered so that the antibody more closely resembles a human antibody, and still retains its original binding ability.

"Hybridization" refers to any process by which a strand of nucleic acid binds with a complementary strand through base pairing.

The term "hybridization complex" refers to a complex formed between two nucleic acid sequences by virtue of the formation of hydrogen bonds between complementary bases. A hybridization complex may be formed in solution (e.g., C_0t or R_0t analysis) or formed between one nucleic acid sequence present in solution and another nucleic acid sequence immobilized on a solid support (e.g., paper, membranes, filters, chips, pins or glass slides, or any other appropriate substrate to which cells or their nucleic acids have been fixed).

The words "insertion" or "addition" refer to changes in an amino acid or nucleotide sequence resulting in the addition of one or more amino acid residues or nucleotides, respectively, to the sequence found in the naturally occurring molecule.

"Immune response" can refer to conditions associated with inflammation, trauma, immune disorders, or infectious or genetic disease, etc. These conditions can be characterized by expression of various factors, e.g., cytokines, chemokines, and other signaling molecules, which may affect cellular and systemic defense systems.

The term "microarray" refers to an arrangement of distinct polynucleotides on a substrate.

The terms "element" or "array element" in a microarray context, refer to hybridizable polynucleotides arranged on the surface of a substrate.

The term "modulate" refers to a change in the activity of HGPRP. For example, modulation may cause an increase or a decrease in protein activity, binding characteristics, or any other biological, functional, or immunological properties of HGPRP.

The phrases "nucleic acid" or "nucleic acid sequence," as used herein, refer to a nucleotide, oligonucleotide, polynucleotide, or any fragment thereof. These phrases also refer to DNA or RNA of genomic or synthetic origin which may be single-stranded or double-stranded and may represent the sense or the antisense strand, to peptide nucleic acid (PNA), or to any DNA-like or RNA-like material. In this context, "fragments" refers to those nucleic acid sequences which comprise a region of unique polynucleotide sequence that specifically identifies SEQ ID NO:7-12, for example, as distinct from any other sequence in the same genome. For example, a fragment of SEQ ID NO:7-12 is useful in hybridization and amplification technologies and in analogous methods that distinguish SEQ ID NO:7-12 from related polynucleotide

sequences. A fragment of SEQ ID NO:7-12 is at least about 15-20 nucleotides in length. The precise length of the fragment of SEQ ID NO:7-12 and the region of SEQ ID NO:7-12 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment. In some cases, a fragment, when translated, would produce polypeptides retaining some functional characteristic, e.g., antigenicity, or structural domain characteristic, e.g., ATP-binding site, of the full-length polypeptide.

The terms "operably associated" or "operably linked" refer to functionally related nucleic acid sequences. A promoter is operably associated or operably linked with a coding sequence if the promoter controls the translation of the encoded polypeptide. While operably associated or operably linked nucleic acid sequences can be contiguous and in the same reading frame, certain genetic elements, e.g., repressor genes, are not contiguously linked to the sequence encoding the polypeptide but still bind to operator sequences that control expression of the polypeptide.

The term "oligonucleotide" refers to a nucleic acid sequence of at least about 6 nucleotides to 60 nucleotides, preferably about 15 to 30 nucleotides, and most preferably about 20 to 25 nucleotides, which can be used in PCR amplification or in a hybridization assay or microarray. "Oligonucleotide" is substantially equivalent to the terms "amplimer," "primer," "oligomer," and "probe," as these terms are commonly defined in the art.

"Peptide nucleic acid" (PNA) refers to an antisense molecule or anti-gene agent which comprises an oligonucleotide of at least about 5 nucleotides in length linked to a peptide backbone of amino acid residues ending in lysine. The terminal lysine confers solubility to the composition. PNAs preferentially bind complementary single stranded DNA or RNA and stop transcript elongation, and may be pegylated to extend their lifespan in the cell.

The term "sample" is used in its broadest sense. A sample suspected of containing nucleic acids encoding HGPRP, or fragments thereof, or HGPRP itself, may comprise a bodily fluid; an extract from a cell, chromosome, organelle, or membrane isolated from a cell; a cell; genomic DNA, RNA, or cDNA, in solution or bound to a substrate; a tissue; a tissue print; etc.

The terms "specific binding" or "specifically binding" refer to that interaction between a protein or peptide and an agonist, an antibody, or an antagonist. The interaction is dependent upon the presence of a particular structure of the protein, e.g., the antigenic determinant or epitope, recognized by the binding molecule. For example, if an antibody is specific for epitope "A," the presence of a polypeptide containing the epitope A, or the presence of free unlabeled A, in a reaction containing free labeled A and the antibody will reduce the amount of labeled A that binds to the antibody.

The term "stringent conditions" refers to conditions which permit hybridization between

polynucleotides and the claimed polynucleotides. Stringent conditions can be defined by salt concentration, the concentration of organic solvent, e.g., formamide, temperature, and other conditions well known in the art. In particular, stringency can be increased by reducing the concentration of salt, increasing the concentration of formamide, or raising the hybridization temperature.

The term "substantially purified" refers to nucleic acid or amino acid sequences that are removed from their natural environment and are isolated or separated, and are at least about 60% free, preferably about 75% free, and most preferably about 90% free from other components with which they are naturally associated.

A "substitution" refers to the replacement of one or more amino acids or nucleotides by different amino acids or nucleotides, respectively.

"Substrate" refers to any suitable rigid or semi-rigid support including membranes, filters, chips, slides, wafers, fibers, magnetic or nonmagnetic beads, gels, tubing, plates, polymers, microparticles and capillaries. The substrate can have a variety of surface forms, such as wells, trenches, pins, channels and pores, to which polynucleotides or polypeptides are bound.

"Transformation" describes a process by which exogenous DNA enters and changes a recipient cell. Transformation may occur under natural or artificial conditions according to various methods well known in the art, and may rely on any known method for the insertion of foreign nucleic acid sequences into a prokaryotic or eukaryotic host cell. The method for transformation is selected based on the type of host cell being transformed and may include, but is not limited to, viral infection, electroporation, heat shock, lipofection, and particle bombardment. The term "transformed" cells includes stably transformed cells in which the inserted DNA is capable of replication either as an autonomously replicating plasmid or as part of the host chromosome, as well as transiently transformed cells which express the inserted DNA or RNA for limited periods of time.

A "variant" of HGPRP polypeptides refers to an amino acid sequence that is altered by one or more amino acid residues. The variant may have "conservative" changes, wherein a substituted amino acid has similar structural or chemical properties (e.g., replacement of leucine with isoleucine). More rarely, a variant may have "nonconservative" changes (e.g., replacement of glycine with tryptophan). Analogous minor variations may also include amino acid deletions or insertions, or both. Guidance in determining which amino acid residues may be substituted, inserted, or deleted without abolishing biological or immunological activity may be found using computer programs well known in the art, for example, LASERGENE software (DNASTAR).

The term "variant", when used in the context of a polynucleotide sequence, may

encompass a polynucleotide sequence related to HGPRP. This definition may also include, for example, "allelic" (as defined above), "splice," "species," or "polymorphic" variants. A splice variant may have significant identity to a reference molecule, but will generally have a greater or lesser number of polynucleotides due to alternate splicing of exons during mRNA processing. The corresponding polypeptide may possess additional functional domains or an absence of domains. Species variants are polynucleotide sequences that vary from one species to another. The resulting polypeptides generally will have significant amino acid identity relative to each other. A polymorphic variant is a variation in the polynucleotide sequence of a particular gene between individuals of a given species. Polymorphic variants also may encompass "single nucleotide polymorphisms" (SNPs) in which the polynucleotide sequence varies by one base. The presence of SNPs may be indicative of, for example, a certain population, a disease state, or a propensity for a disease state.

THE INVENTION

The invention is based on the discovery of new human GPCR proteins (HGPRP), the polynucleotides encoding HGPRP, and the use of these compositions for the diagnosis, treatment, or prevention of cell proliferative, neurological, and immune disorders.

Table 1 lists the Incyte Clones used to derive full length nucleotide sequences encoding HGPRP. Columns 1 and 2 show the sequence identification numbers (SEQ ID NO) of the amino acid and nucleic acid sequences, respectively. Column 3 shows the Clone ID of the Incyte Clone in which nucleic acids encoding each HGPRP were identified, and column 4, the cDNA libraries from which these clones were isolated. Column 5 shows Incyte clones, their corresponding cDNA libraries, and shotgun sequences. The clones and shotgun sequences are part of the consensus nucleotide sequence of each HGPRP and are useful as fragments in hybridization technologies.

The columns of Table 2 show various properties of the polypeptides of the invention: column 1 references the SEQ ID NO; column 2 shows the number of amino acid residues in each polypeptide; column 3, potential phosphorylation sites; column 4, potential glycosylation sites; column 5, the amino acid residues comprising signature sequences and motifs; column 6, the identity of each protein; and column 7, analytical methods used to identify each protein through sequence homology and protein motifs.

The columns of Table 3 show the tissue-specificity and diseases, disorders, or conditions associated with nucleotide sequences encoding HGPRP. The first column of Table 3 lists the polypeptide sequence identifiers. The second column lists tissue categories which express HGPRP as a fraction of total tissue categories expressing HGPRP. The third column lists the diseases, disorders, or conditions associated with those tissues expressing HGPRP. The fourth column lists

the vectors used to subclone the cDNA library.

The following fragments of the nucleotide sequences encoding HGPRP are useful in hybridization or amplification technologies to identify SEQ ID NO:7-12 and to distinguish between SEQ ID NO:7-12 and related polynucleotide sequences. The useful fragments are the fragment of SEQ ID NO:7 from about nucleotide 235 to about nucleotide 270; the fragment of SEQ ID NO:8 from about nucleotide 218 to about nucleotide 247; the fragment of SEQ ID NO:9 from about nucleotide 271 to about nucleotide 300; the fragment of SEQ ID NO:10 from about nucleotide 273 to about nucleotide 303; the fragment of SEQ ID NO:11 from about nucleotide 542 to about nucleotide 571; and the fragment of SEQ ID NO:12 from about nucleotide 703 to about nucleotide 735.

The invention also encompasses HGPRP variants. A preferred HGPRP variant is one which has at least about 80%, more preferably at least about 90%, and most preferably at least about 95% amino acid sequence identity to the HGPRP amino acid sequence, and which contains at least one functional or structural characteristic of HGPRP.

The invention also encompasses polynucleotides which encode HGPRP. In a particular embodiment, the invention encompasses a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:7-12, which encodes HGPRP.

The invention also encompasses a variant of a polynucleotide sequence encoding HGPRP. In particular, such a variant polynucleotide sequence will have at least about 70%, more preferably at least about 85%, and most preferably at least about 95% polynucleotide sequence identity to the polynucleotide sequence encoding HGPRP. A particular aspect of the invention encompasses a variant of a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:7-12 which has at least about 70%. more preferably at least about 85%, and most preferably at least about 95% polynucleotide sequence identity to a nucleic acid sequence selected from the group consisting of SEQ ID NO:7-12. Any one of the polynucleotide variants described above can encode an amino acid sequence which contains at least one functional or structural characteristic of HGPRP.

It will be appreciated by those skilled in the art that, as a result of the degeneracy of the genetic code, a multitude of polynucleotide sequences encoding HGPRP, some bearing minimal similarity to the polynucleotide sequences of any known and naturally occurring gene, may be produced. Thus, the invention contemplates each and every possible variation of polynucleotide sequence that could be made by selecting combinations based on possible codon choices. These combinations are made in accordance with the standard triplet genetic code as applied to the polynucleotide sequence of naturally occurring HGPRP, and all such variations are to be

considered as being specifically disclosed.

Although nucleotide sequences which encode HGPRP and its variants are preferably capable of hybridizing to the nucleotide sequence of the naturally occurring HGPRP under appropriately selected conditions of stringency, it may be advantageous to produce nucleotide sequences encoding HGPRP or its derivatives possessing a substantially different codon usage, e.g., inclusion of non-naturally occurring codons. Codons may be selected to increase the rate at which expression of the peptide occurs in a particular prokaryotic or eukaryotic host in accordance with the frequency with which particular codons are utilized by the host. Other reasons for substantially altering the nucleotide sequence encoding HGPRP and its derivatives without altering the encoded amino acid sequences include the production of RNA transcripts having more desirable properties, such as a greater half-life, than transcripts produced from the naturally occurring sequence.

The invention also encompasses production of DNA sequences which encode HGPRP and HGPRP derivatives, or fragments thereof, entirely by synthetic chemistry. After production, the synthetic sequence may be inserted into any of the many available expression vectors and cell systems using reagents well known in the art. Moreover, synthetic chemistry may be used to introduce mutations into a sequence encoding HGPRP or any fragment thereof.

Also encompassed by the invention are polynucleotide sequences that are capable of hybridizing to the claimed polynucleotide sequences, and, in particular, to those shown in SEQ ID NO:7-12 and fragments thereof under various conditions of stringency. (See, e.g., Wahl, G.M. and S.L. Berger (1987) Methods Enzymol. 152:399-407; Kimmel, A.R. (1987) Methods Enzymol. 152:507-511.) For example, stringent salt concentration will ordinarily be less than about 750 mM NaCl and 75 mM trisodium citrate, preferably less than about 500 mM NaCl and 50 mM trisodium citrate, and most preferably less than about 250 mM NaCl and 25 mM trisodium citrate. Low stringency hybridization can be obtained in the absence of organic solvent, e.g., formamide, while high stringency hybridization can be obtained in the presence of at least about 35% formamide, and most preferably at least about 50% formamide. Stringent temperature conditions will ordinarily include temperatures of at least about 30°C, more preferably of at least about 37°C, and most preferably of at least about 42°C. Varying additional parameters, such as hybridization time, the concentration of detergent, e.g., sodium dodecyl sulfate (SDS), and the inclusion or exclusion of carrier DNA, are well known to those skilled in the art. Various levels of stringency are accomplished by combining these various conditions as needed. In a preferred embodiment, hybridization will occur at 30°C in 750 mM NaCl, 75 mM trisodium citrate, and 1% SDS. In a more preferred embodiment, hybridization will occur at 37°C in 500 mM NaCl, 50 mM trisodium

citrate, 1% SDS, 35% formamide, and 100 μg/ml denatured salmon sperm DNA (ssDNA). In a most preferred embodiment, hybridization will occur at 42°C in 250 mM NaCl, 25 mM trisodium citrate, 1% SDS, 50 % formamide, and 200 μg/ml ssDNA. Useful variations on these conditions will be readily apparent to those skilled in the art.

The washing steps which follow hybridization can also vary in stringency. Wash stringency conditions can be defined by salt concentration and by temperature. As above, wash stringency can be increased by decreasing salt concentration or by increasing temperature. For example, stringent salt concentration for the wash steps will preferably be less than about 30 mM NaCl and 3 mM trisodium citrate, and most preferably less than about 15 mM NaCl and 1.5 mM trisodium citrate. Stringent temperature conditions for the wash steps will ordinarily include temperature of at least about 25°C, more preferably of at least about 42°C, and most preferably of at least about 68°C. In a preferred embodiment, wash steps will occur at 25°C in 30 mM NaCl, 3 mM trisodium citrate, and 0.1% SDS. In a more preferred embodiment, wash steps will occur at 42°C in 15 mM NaCl, 1.5 mM trisodium citrate, and 0.1% SDS. In a most preferred embodiment, wash steps will occur at 68°C in 15 mM NaCl, 1.5 mM trisodium citrate, and 0.1% SDS. Additional variations on these conditions will be readily apparent to those skilled in the art.

Methods for DNA sequencing are well known in the art and may be used to practice any of the embodiments of the invention. The methods may employ such enzymes as the Klenow fragment of DNA polymerase I, SEQUENASE (US Biochemical, Cleveland OH), Taq DNA polymerase (PE Biosystems), thermostable T7 polymerase (Amersham Pharmacia Biotech, Piscataway NJ), or combinations of polymerases and proofreading exonucleases such as those found in the ELONGASE amplification system (Life Technologies, Gaithersburg MD). Preferably, sequence preparation is automated with machines such as the MICROLAB 2200 system (Hamilton, Reno NV), DNA ENGINE thermal cycler (PTC200; MJ Research, Watertown MA) and the ABI CATALYST 800 (PE Biosystems). Sequencing is then carried out using either ABI PRISM 373 or 377 DNA sequencing systems (PE Biosystems) or the MEGABACE 1000 DNA sequencing system (Amersham Pharmacia Biotech). The resulting sequences are analyzed using a variety of algorithms which are well known in the art. (See, e.g., Ausubel, F.M. (1997) Short Protocols in Molecular Biology, John Wiley & Sons, New York NY, unit 7.7; Meyers, R.A. (1995) Molecular Biology and Biotechnology, Wiley VCH, New York NY, pp. 856-853.)

The nucleic acid sequences encoding HGPRP may be extended utilizing a partial nucleotide sequence and employing various PCR-based methods known in the art to detect upstream sequences, such as promoters and regulatory elements. For example, one method which may be employed, restriction-site PCR, uses universal and nested primers to amplify unknown

sequence from genomic DNA within a cloning vector. (See, e.g., Sarkar, G. (1993) PCR Methods Applic. 2:318-322.) Another method, inverse PCR, uses primers that extend in divergent directions to amplify unknown sequence from a circularized template. The template is derived from restriction fragments comprising a known genomic locus and surrounding sequences. (See, e.g., Triglia, T. et al. (1988) Nucleic Acids Res. 16:8186.) A third method, capture PCR, involves PCR amplification of DNA fragments adjacent to known sequences in human and yeast artificial chromosome DNA. (See, e.g., Lagerstrom, M. et al. (1991) PCR Methods Applic. 1:111-119.) In this method, multiple restriction enzyme digestions and ligations may be used to insert an engineered double-stranded sequence into a region of unknown sequence before performing PCR. Other methods which may be used to retrieve unknown sequences are known in the art. (See, e.g., Parker, J.D. et al. (1991) Nucleic Acids Res. 19:3055-3060). Additionally, one may use PCR, nested primers, and PROMOTERFINDER libraries (Clontech, Palo Alto CA) to walk genomic DNA. This procedure avoids the need to screen libraries and is useful in finding intron/exon junctions. For all PCR-based methods, primers may be designed using commercially available software, such as OLIGO 4.06 primer analysis software (National Biosciences, Plymouth MN) or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the template at temperatures of about 68°C to 72°C.

When screening for full-length cDNAs, it is preferable to use libraries that have been size-selected to include larger cDNAs. In addition, random-primed libraries, which often include sequences containing the 5' regions of genes, are preferable for situations in which an oligo d(T) library does not yield a full-length cDNA. Genomic libraries may be useful for extension of sequence into 5' non-transcribed regulatory regions.

Capillary electrophoresis systems which are commercially available may be used to analyze the size or confirm the nucleotide sequence of sequencing or PCR products. In particular, capillary sequencing may employ flowable polymers for electrophoretic separation, four different nucleotide-specific, laser-stimulated fluorescent dyes, and a charge coupled device camera for detection of the emitted wavelengths. Output/light intensity may be converted to electrical signal using appropriate software (e.g., GENOTYPER and SEQUENCE NAVIGATOR software, PE Biosystems), and the entire process from loading of samples to computer analysis and electronic data display may be computer controlled. Capillary electrophoresis is especially preferable for sequencing small DNA fragments which may be present in limited amounts in a particular sample.

In another embodiment of the invention, polynucleotide sequences or fragments thereof which encode HGPRP may be cloned in recombinant DNA molecules that direct expression of HGPRP, or fragments or functional equivalents thereof, in appropriate host cells. Due to the

inherent degeneracy of the genetic code, other DNA sequences which encode substantially the same or a functionally equivalent amino acid sequence may be produced and used to express HGPRP.

The nucleotide sequences of the present invention can be engineered using methods generally known in the art in order to alter HGPRP-encoding sequences for a variety of purposes including, but not limited to, modification of the cloning, processing, and/or expression of the gene product. DNA shuffling by random fragmentation and PCR reassembly of gene fragments and synthetic oligonucleotides may be used to engineer the nucleotide sequences. For example, oligonucleotide-mediated site-directed mutagenesis may be used to introduce mutations that create new restriction sites, alter glycosylation patterns, change codon preference, produce splice variants, and so forth.

In another embodiment, sequences encoding HGPRP may be synthesized, in whole or in part, using chemical methods well known in the art. (See, e.g., Caruthers, M.H. et al. (1980) Nucleic Acids Symp. Ser. 7:215-223; Horn, T. et al. (1980) Nucleic Acids Symp. Ser. 7:225-232.) Alternatively, HGPRP itself or a fragment thereof may be synthesized using chemical methods. For example, peptide synthesis can be performed using various solid-phase techniques. (See, e.g., Roberge, J.Y. et al. (1995) Science 269:202-204.) Automated synthesis may be achieved using the ABI 431A peptide synthesizer (PE Biosystems). Additionally, the amino acid sequence of HGPRP, or any part thereof, may be altered during direct synthesis and/or combined with sequences from other proteins, or any part thereof, to produce a variant polypeptide.

The peptide may be substantially purified by preparative high performance liquid chromatography. (See, e.g, Chiez, R.M. and F.Z. Regnier (1990) Methods Enzymol. 182:392-421.) The composition of the synthetic peptides may be confirmed by amino acid analysis or by sequencing. (See, e.g., Creighton, T. (1984) Proteins, Structures and Molecular Properties, WH Freeman, New York NY.)

In order to express a biologically active HGPRP, the nucleotide sequences encoding HGPRP or derivatives thereof may be inserted into an appropriate expression vector, i.e., a vector which contains the necessary elements for transcriptional and translational control of the inserted coding sequence in a suitable host. These elements include regulatory sequences, such as enhancers, constitutive and inducible promoters, and 5' and 3' untranslated regions in the vector and in polynucleotide sequences encoding HGPRP. Such elements may vary in their strength and specificity. Specific initiation signals may also be used to achieve more efficient translation of sequences encoding HGPRP. Such signals include the ATG initiation codon and adjacent sequences, e.g. the Kozak sequence. In cases where sequences encoding HGPRP and its initiation

codon and upstream regulatory sequences are inserted into the appropriate expression vector, no additional transcriptional or translational control signals may be needed. However, in cases where only coding sequence, or a fragment thereof, is inserted, exogenous translational control signals including an in-frame ATG initiation codon should be provided by the vector. Exogenous translational elements and initiation codons may be of various origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of enhancers appropriate for the particular host cell system used. (See, e.g., Scharf, D. et al. (1994) Results Probl. Cell Differ. 20:125-162.)

Methods which are well known to those skilled in the art may be used to construct expression vectors containing sequences encoding HGPRP and appropriate transcriptional and translational control elements. These methods include in vitro recombinant DNA techniques, synthetic techniques, and in vivo genetic recombination. (See, e.g., Sambrook, J. et al. (1989) Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press, Plainview NY, ch. 4, 8, and 16-17; Ausubel, F.M. et al. (1995) Current Protocols in Molecular Biology, John Wiley & Sons, New York NY, ch. 9, 13, and 16.)

A variety of expression vector/host systems may be utilized to contain and express sequences encoding HGPRP. These include, but are not limited to, microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid, or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with viral expression vectors (e.g., baculovirus); plant cell systems transformed with viral expression vectors (e.g., cauliflower mosaic virus, CaMV, or tobacco mosaic virus,TMV) or with bacterial expression vectors (e.g., Ti or pBR322 plasmids); or animal cell systems. The invention is not limited by the host cell employed.

In bacterial systems, a number of cloning and expression vectors may be selected depending upon the use intended for polynucleotide sequences encoding HGPRP. For example, routine cloning, subcloning, and propagation of polynucleotide sequences encoding HGPRP can be achieved using a multifunctional E. coli vector such as PBLUESCRIPT (Stratagene, La Jolla CA) or PSPORT1 plasmid (Life Technologies). Ligation of sequences encoding HGPRP into the vector's multiple cloning site disrupts the *lacZ* gene, allowing a colorimetric screening procedure for identification of transformed bacteria containing recombinant molecules. In addition, these vectors may be useful for in vitro transcription, dideoxy sequencing, single strand rescue with helper phage, and creation of nested deletions in the cloned sequence. (See, e.g., Van Heeke, G. and S.M. Schuster (1989) J. Biol. Chem. 264:5503-5509.) When large quantities of HGPRP are needed, e.g. for the production of antibodies, vectors which direct high level expression of HGPRP

may be used. For example, vectors containing the strong, inducible T5 or T7 bacteriophage promoter may be used.

Yeast expression systems may be used for production of HGPRP. A number of vectors containing constitutive or inducible promoters, such as alpha factor, alcohol oxidase, and PGH, may be used in the yeast Saccharomyces cerevisiae or Pichia pastoris. In addition, such vectors direct either the secretion or intracellular retention of expressed proteins and enable integration of foreign sequences into the host genome for stable propagation. (See, e.g., Ausubel, 1995, supra; Bitter, G.A. et al. (1987) Methods Enzymol. 153:516-544; and Scorer, C.A. et al. (1994) Bio/Technology 12:181-184.)

Plant systems may also be used for expression of HGPRP. Transcription of sequences encoding HGPRP may be driven viral promoters, e.g., the 35S and 19S promoters of CaMV used alone or in combination with the omega leader sequence from TMV (Takamatsu, N. (1987) EMBO J. 6:307-311). Alternatively, plant promoters such as the small subunit of RUBISCO or heat shock promoters may be used. (See, e.g., Coruzzi, G. et al. (1984) EMBO J. 3:1671-1680; Broglie, R. et al. (1984) Science 224:838-843; and Winter, J. et al. (1991) Results Probl. Cell Differ. 17:85-105.) These constructs can be introduced into plant cells by direct DNA transformation or pathogen-mediated transfection. (See, e.g., The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196.)

In mammalian cells, a number of viral-based expression systems may be utilized. In cases where an adenovirus is used as an expression vector, sequences encoding HGPRP may be ligated into an adenovirus transcription/translation complex consisting of the late promoter and tripartite leader sequence. Insertion in a non-essential E1 or E3 region of the viral genome may be used to obtain infective virus which expresses HGPRP in host cells. (See, e.g., Logan, J. and T. Shenk (1984) Proc. Natl. Acad. Sci. USA 81:3655-3659.) In addition, transcription enhancers, such as the Rous sarcoma virus (RSV) enhancer, may be used to increase expression in mammalian host cells. SV40 or EBV-based vectors may also be used for high-level protein expression.

Human artificial chromosomes (HACs) may also be employed to deliver larger fragments of DNA than can be contained in and expressed from a plasmid. HACs of about 6 kb to 10 Mb are constructed and delivered via conventional delivery methods (liposomes, polycationic amino polymers, or vesicles) for therapeutic purposes. (See, e.g., Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355.)

For long term production of recombinant proteins in mammalian systems, stable expression of HGPRP in cell lines is preferred. For example, sequences encoding HGPRP can be transformed into cell lines using expression vectors which may contain viral origins of replication

and/or endogenous expression elements and a selectable marker gene on the same or on a separate vector. Following the introduction of the vector, cells may be allowed to grow for about 1 to 2 days in enriched media before being switched to selective media. The purpose of the selectable marker is to confer resistance to a selective agent, and its presence allows growth and recovery of cells which successfully express the introduced sequences. Resistant clones of stably transformed cells may be propagated using tissue culture techniques appropriate to the cell type.

Any number of selection systems may be used to recover transformed cell lines. These include, but are not limited to, the herpes simplex virus thymidine kinase and adenine phosphoribosyltransferase genes, for use in *tk*⁻ or *apr*⁻ cells, respectively. (See, e.g., Wigler, M. et al. (1977) Cell 11:223-232; Lowy, I. et al. (1980) Cell 22:817-823.) Also, antimetabolite, antibiotic, or herbicide resistance can be used as the basis for selection. For example, *dhfr* confers resistance to methotrexate; *neo* confers resistance to the aminoglycosides, neomycin and G-418; and *als* and *pat* confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively. (See, e.g., Wigler, M. et al. (1980) Proc. Natl. Acad. Sci. USA 77:3567-3570; Colbere-Garapin, F. et al. (1981) J. Mol. Biol. 150:1-14.) Additional selectable genes have been described, e.g., *trpB* and *hisD*, which alter cellular requirements for metabolites. (See, e.g., Hartman, S.C. and R.C. Mulligan (1988) Proc. Natl. Acad. Sci. USA 85:8047-8051.) Visible markers, e.g., anthocyanins, green fluorescent proteins (GFP; Clontech), ß glucuronidase and its substrate ß-glucuronide, or luciferase and its substrate luciferin may be used. These markers can be used not only to identify transformants, but also to quantify the amount of transient or stable protein expression attributable to a specific vector system. (See, e.g., Rhodes, C.A. (1995) Methods Mol. Biol. 55:121-131.)

Although the presence/absence of marker gene expression suggests that the gene of interest is also present, the presence and expression of the gene may need to be confirmed. For example, if the sequence encoding HGPRP is inserted within a marker gene sequence, transformed cells containing sequences encoding HGPRP can be identified by the absence of marker gene function. Alternatively, a marker gene can be placed in tandem with a sequence encoding HGPRP under the control of a single promoter. Expression of the marker gene in response to induction or selection usually indicates expression of the tandem gene as well.

In general, host cells that contain the nucleic acid sequence encoding HGPRP and that express HGPRP may be identified by a variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridizations, PCR amplification, and protein bioassay or immunoassay techniques which include membrane, solution, or chip based technologies for the detection and/or quantification of nucleic acid or

protein sequences.

Immunological methods for detecting and measuring the expression of HGPRP using either specific polyclonal or monoclonal antibodies are known in the art. Examples of such techniques include enzyme-linked immunosorbent assays (ELISAs), radioimmunoassays (RIAs), and fluorescence activated cell sorting (FACS). A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering epitopes on HGPRP is preferred, but a competitive binding assay may be employed. These and other assays are well known in the art. (See, e.g., Hampton, R. et al. (1990) Serological Methods, a Laboratory Manual, APS Press, St. Paul MN, Sect. IV; Coligan, J.E. et al. (1997) Current Protocols in Immunology, Greene Pub. Associates and Wiley-Interscience, New York NY; and Pound, J.D. (1998) Immunochemical Protocols, Humana Press, Totowa NJ.)

A wide variety of labels and conjugation techniques are known by those skilled in the art and may be used in various nucleic acid and amino acid assays. Means for producing labeled hybridization or PCR probes for detecting sequences related to polynucleotides encoding HGPRP include oligolabeling, nick translation, end-labeling, or PCR amplification using a labeled nucleotide. Alternatively, the sequences encoding HGPRP, or any fragments thereof, may be cloned into a vector for the production of an mRNA probe. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by addition of an appropriate RNA polymerase such as T7, T3, or SP6 and labeled nucleotides. These procedures may be conducted using a variety of commercially available kits, such as those provided by Amersham Pharmacia Biotech, Promega (Madison WI), and US Biochemical. Suitable reporter molecules or labels which may be used for ease of detection include radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents, as well as substrates, cofactors, inhibitors, magnetic particles, and the like.

Host cells transformed with nucleotide sequences encoding HGPRP may be cultured under conditions suitable for the expression and recovery of the protein from cell culture. The protein produced by a transformed cell may be secreted or retained intracellularly depending on the sequence and/or the vector used. As will be understood by those of skill in the art, expression vectors containing polynucleotides which encode HGPRP may be designed to contain signal sequences which direct secretion of HGPRP through a prokaryotic or eukaryotic cell membrane.

In addition, a host cell strain may be chosen for its ability to modulate expression of the inserted sequences or to process the expressed protein in the desired fashion. Such modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation, and acylation. Post-translational processing which cleaves a "prepro"

form of the protein may also be used to specify protein targeting, folding, and/or activity. Different host cells which have specific cellular machinery and characteristic mechanisms for post-translational activities (e.g., CHO, HeLa, MDCK, HEK293, and WI38), are available from the American Type Culture Collection (ATCC, Manassas VA) and may be chosen to ensure the correct modification and processing of the foreign protein.

In another embodiment of the invention, natural, modified, or recombinant nucleic acid sequences encoding HGPRP may be ligated to a heterologous sequence resulting in translation of a fusion protein in any of the aforementioned host systems. For example, a chimeric HGPRP protein containing a heterologous moiety that can be recognized by a commercially available antibody may facilitate the screening of peptide libraries for inhibitors of HGPRP activity. Heterologous protein and peptide moieties may also facilitate purification of fusion proteins using commercially available affinity matrices. Such moieties include, but are not limited to, glutathione S-transferase (GST), maltose binding protein (MBP), thioredoxin (Trx), calmodulin binding peptide (CBP), 6-His, FLAG, *c-myc*, and hemagglutinin (HA). GST, MBP, Trx, CBP, and 6-His enable purification of their cognate fusion proteins on immobilized glutathione, maltose, phenylarsine oxide, calmodulin, and metal-chelate resins, respectively. FLAG, *c-myc*, and hemagglutinin (HA) enable immunoaffinity purification of fusion proteins using commercially available monoclonal and polyclonal antibodies that specifically recognize these epitope tags. A fusion protein may also be engineered to contain a proteolytic cleavage site located between the HGPRP encoding sequence and the heterologous protein sequence, so that HGPRP may be cleaved away from the heterologous moiety following purification. Methods for fusion protein expression and purification are discussed in Ausubel (1995, supra, ch 10). A variety of commercially available kits may also be used to facilitate expression and purification of fusion proteins.

In a further embodiment of the invention, synthesis of radiolabeled HGPRP may be achieved in vitro using the TNT rabbit reticulocyte lysate or wheat germ extract systems (Promega). These systems couple transcription and translation of protein-coding sequences operably associated with the T7, T3, or SP6 promoters. Translation takes place in the presence of a radiolabeled amino acid precursor, preferably ^{35}S-methionine.

Fragments of HGPRP may be produced not only by recombinant production, but also by direct peptide synthesis using solid-phase techniques. (See, e.g., Creighton, supra, pp. 55-60.) Protein synthesis may be performed by manual techniques or by automation. Automated synthesis may be achieved, for example, using the ABI 431A peptide synthesizer (PE Biosystems). Various fragments of HGPRP may be synthesized separately and then combined to produce the full length

molecule.

THERAPEUTICS

Chemical and structural similarity, e.g., in the context of sequences and motifs, exists between regions of HGPRP and GPCR proteins. In addition, the expression of HGPRP is closely associated with cell proliferative and immune disorders, and with neurological tissues. Therefore, HGPRP appears to play a role in cell proliferative, neurological, and immune disorders. In the treatment of disorders associated with increased HGPRP expression or activity, it is desirable to decrease the expression or activity of HGPRP. In the treatment of disorders associated with decreased HGPRP expression or activity, it is desirable to increase the expression or activity of HGPRP.

Therefore, in one embodiment, HGPRP or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of HGPRP. Examples of such disorders include, but are not limited to, a cell proliferative disorder such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia; cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus; an immune disorder such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; and a neurological disorder such as epilepsy, ischemic cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer's disease, Pick's disease,

Huntington's disease, dementia, Parkinson's disease and other extrapyramidal disorders, amyotrophic lateral sclerosis and other motor neuron disorders, progressive neural muscular atrophy, retinitis pigmentosa, hereditary ataxias, multiple sclerosis and other demyelinating diseases, bacterial and viral meningitis, brain abscess, subdural empyema, epidural abscess, suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral central nervous system disease; prion diseases including kuru, Creutzfeldt-Jakob disease, and Gerstmann-Straussler-Scheinker syndrome; fatal familial insomnia, nutritional and metabolic diseases of the nervous system, neurofibromatosis, tuberous sclerosis, cerebelloretinal hemangioblastomatosis, encephalotrigeminal syndrome, mental retardation and other developmental disorders of the central nervous system, cerebral palsy, neuroskeletal disorders, autonomic nervous system disorders, cranial nerve disorders, spinal cord diseases, muscular dystrophy and other neuromuscular disorders, peripheral nervous system disorders, dermatomyositis and polymyositis; inherited, metabolic, endocrine, and toxic myopathies; myasthenia gravis, periodic paralysis; mental disorders including mood, anxiety, and schizophrenic disorders; akathesia, amnesia, catatonia, diabetic neuropathy, tardive dyskinesia, dystonias, paranoid psychoses, postherpetic neuralgia, and Tourette's disorder.

In another embodiment, a vector capable of expressing HGPRP or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of HGPRP including, but not limited to, those described above.

In a further embodiment, a pharmaceutical composition comprising a substantially purified HGPRP in conjunction with a suitable pharmaceutical carrier may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of HGPRP including, but not limited to, those provided above.

In still another embodiment, an agonist which modulates the activity of HGPRP may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of HGPRP including, but not limited to, those listed above.

In a further embodiment, an antagonist of HGPRP may be administered to a subject to treat or prevent a disorder associated increased expression or activity of HGPRP. Examples of such disorders include, but are not limited to, those described above. In one aspect, an antibody which specifically binds HGPRP may be used directly as an antagonist or indirectly as a targeting or delivery mechanism for bringing a pharmaceutical agent to cells or tissue which express HGPRP.

In an additional embodiment, a vector expressing the complement of the polynucleotide encoding HGPRP may be administered to a subject to treat or prevent a disorder associated

increased expression or activity of HGPRP including, but not limited to, those described above.

In other embodiments, any of the proteins, antagonists, antibodies, agonists, complementary sequences, or vectors of the invention may be administered in combination with other appropriate therapeutic agents. Selection of the appropriate agents for use in combination therapy may be made by one of ordinary skill in the art, according to conventional pharmaceutical principles. The combination of therapeutic agents may act synergistically to effect the treatment or prevention of the various disorders described above. Using this approach, one may be able to achieve therapeutic efficacy with lower dosages of each agent, thus reducing the potential for adverse side effects.

An antagonist of HGPRP may be produced using methods which are generally known in the art. In particular, purified HGPRP may be used to produce antibodies or to screen libraries of pharmaceutical agents to identify those which specifically bind HGPRP. Antibodies to HGPRP may also be generated using methods that are well known in the art. Such antibodies may include, but are not limited to, polyclonal, monoclonal, chimeric, and single chain antibodies, Fab fragments, and fragments produced by a Fab expression library. Neutralizing antibodies (e.g., those which inhibit dimer formation) are especially preferred for therapeutic use.

For the production of antibodies, various hosts including goats, rabbits, rats, mice, humans, and others may be immunized by injection with HGPRP or with any fragment or oligopeptide thereof which has immunogenic properties. Depending on the host species, various adjuvants may be used to increase immunological response. Such adjuvants include, but are not limited to, Freund's, mineral gels such as aluminum hydroxide, and surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, KLH, and dinitrophenol. Among adjuvants used in humans, BCG (bacilli Calmette-Guerin) and Corynebacterium parvum are especially preferable.

It is preferred that the oligopeptides, peptides, or fragments used to induce antibodies to HGPRP have an amino acid sequence consisting of at least about 5 amino acids, and, more preferably, of at least about 10 amino acids. It is also preferable that these oligopeptides, peptides, or fragments are identical to a portion of the amino acid sequence of the natural protein and contain the entire amino acid sequence of a small, naturally occurring molecule. Short stretches of HGPRP amino acids may be fused with those of another protein, such as KLH, and antibodies to the chimeric molecule may be produced.

Monoclonal antibodies to HGPRP may be prepared using any technique which provides for the production of antibody molecules by continuous cell lines in culture. These include, but are not limited to, the hybridoma technique, the human B-cell hybridoma technique, and the EBV-

hybridoma technique. (See, e.g., Kohler, G. et al. (1975) Nature 256:495-497; Kozbor, D. et al. (1985) J. Immunol. Methods 81:31-42; Cote, R.J. et al. (1983) Proc. Natl. Acad. Sci. USA 80:2026-2030; and Cole, S.P. et al. (1984) Mol. Cell Biol. 62:109-120.)

In addition, techniques developed for the production of "chimeric antibodies," such as the splicing of mouse antibody genes to human antibody genes to obtain a molecule with appropriate antigen specificity and biological activity, can be used. (See, e.g., Morrison, S.L. et al. (1984) Proc. Natl. Acad. Sci. USA 81:6851-6855; Neuberger, M.S. et al. (1984) Nature 312:604-608; and Takeda, S. et al. (1985) Nature 314:452-454.) Alternatively, techniques described for the production of single chain antibodies may be adapted, using methods known in the art, to produce HGPRP-specific single chain antibodies. Antibodies with related specificity, but of distinct idiotypic composition, may be generated by chain shuffling from random combinatorial immunoglobulin libraries. (See, e.g., Burton, D.R. (1991) Proc. Natl. Acad. Sci. USA 88:10134-10137.)

Antibodies may also be produced by inducing in vivo production in the lymphocyte population or by screening immunoglobulin libraries or panels of highly specific binding reagents as disclosed in the literature. (See, e.g., Orlandi, R. et al. (1989) Proc. Natl. Acad. Sci. USA 86:3833-3837; Winter, G. et al. (1991) Nature 349:293-299.)

Antibody fragments which contain specific binding sites for HGPRP may also be generated. For example, such fragments include, but are not limited to, F(ab')2 fragments produced by pepsin digestion of the antibody molecule and Fab fragments generated by reducing the disulfide bridges of the F(ab')2 fragments. Alternatively, Fab expression libraries may be constructed to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity. (See, e.g., Huse, W.D. et al. (1989) Science 246:1275-1281.)

Various immunoassays may be used for screening to identify antibodies having the desired specificity. Numerous protocols for competitive binding or immunoradiometric assays using either polyclonal or monoclonal antibodies with established specificities are well known in the art. Such immunoassays typically involve the measurement of complex formation between HGPRP and its specific antibody. A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering HGPRP epitopes is preferred, but a competitive binding assay may also be employed (Pound, supra).

Various methods such as Scatchard analysis in conjunction with radioimmunoassay techniques may be used to assess the affinity of antibodies for HGPRP. Affinity is expressed as an association constant, K_a, which is defined as the molar concentration of HGPRP-antibody complex divided by the molar concentrations of free antigen and free antibody under equilibrium conditions. The K_a determined for a preparation of polyclonal antibodies, which are

heterogeneous in their affinities for multiple HGPRP epitopes, represents the average affinity, or avidity, of the antibodies for HGPRP. The K_a determined for a preparation of monoclonal antibodies, which are monospecific for a particular HGPRP epitope, represents a true measure of affinity. High-affinity antibody preparations with K_a ranging from about 10^9 to 10^{12} L/mole are preferred for use in immunoassays in which the HGPRP-antibody complex must withstand rigorous manipulations. Low-affinity antibody preparations with K_a ranging from about 10^6 to 10^7 L/mole are preferred for use in immunopurification and similar procedures which ultimately require dissociation of HGPRP, preferably in active form, from the antibody (Catty, D. (1988) Antibodies, Volume I: A Practical Approach, IRL Press, Washington, DC; Liddell, J.E. and Cryer, A. (1991) A Practical Guide to Monoclonal Antibodies, John Wiley & Sons, New York NY).

The titer and avidity of polyclonal antibody preparations may be further evaluated to determine the quality and suitability of such preparations for certain downstream applications. For example, a polyclonal antibody preparation containing at least 1-2 mg specific antibody/ml, preferably 5-10 mg specific antibody/ml, is preferred for use in procedures requiring precipitation of HGPRP-antibody complexes. Procedures for evaluating antibody specificity, titer, and avidity, and guidelines for antibody quality and usage in various applications, are generally available. (See, e.g., Catty, supra, and Coligan et al. supra.)

In another embodiment of the invention, the polynucleotides encoding HGPRP, or any fragment or complement thereof, may be used for therapeutic purposes. In one aspect, the complement of the polynucleotide encoding HGPRP may be used in situations in which it would be desirable to block the transcription of the mRNA. In particular, cells may be transformed with sequences complementary to polynucleotides encoding HGPRP. Thus, complementary molecules or fragments may be used to modulate HGPRP activity, or to achieve regulation of gene function. Such technology is now well known in the art, and sense or antisense oligonucleotides or larger fragments can be designed from various locations along the coding or control regions of sequences encoding HGPRP.

Expression vectors derived from retroviruses, adenoviruses, or herpes or vaccinia viruses, or from various bacterial plasmids, may be used for delivery of nucleotide sequences to the targeted organ, tissue, or cell population. Methods which are well known to those skilled in the art can be used to construct vectors to express nucleic acid sequences complementary to the polynucleotides encoding HGPRP. (See, e.g., Sambrook, supra; Ausubel, 1995, supra.)

Genes encoding HGPRP can be turned off by transforming a cell or tissue with expression vectors which express high levels of a polynucleotide, or fragment thereof, encoding HGPRP. Such constructs may be used to introduce untranslatable sense or antisense sequences into a cell. Even in the absence of integration into the DNA, such vectors may continue to transcribe RNA

molecules until they are disabled by endogenous nucleases. Transient expression may last for a month or more with a non-replicating vector, and may last even longer if appropriate replication elements are part of the vector system.

As mentioned above, modifications of gene expression can be obtained by designing complementary sequences or antisense molecules (DNA, RNA, or PNA) to the control, 5', or regulatory regions of the gene encoding HGPRP. Oligonucleotides derived from the transcription initiation site, e.g., between about positions -10 and +10 from the start site, are preferred. Similarly, inhibition can be achieved using triple helix base-pairing methodology. Triple helix pairing is useful because it causes inhibition of the ability of the double helix to open sufficiently for the binding of polymerases, transcription factors, or regulatory molecules. Recent therapeutic advances using triplex DNA have been described in the literature. (See, e.g., Gee, J.E. et al. (1994) in Huber, B.E. and B.I. Carr, Molecular and Immunologic Approaches, Futura Publishing, Mt. Kisco NY, pp. 163-177.) A complementary sequence or antisense molecule may also be designed to block translation of mRNA by preventing the transcript from binding to ribosomes.

Ribozymes, enzymatic RNA molecules, may also be used to catalyze the specific cleavage of RNA. The mechanism of ribozyme action involves sequence-specific hybridization of the ribozyme molecule to complementary target RNA, followed by endonucleolytic cleavage. For example, engineered hammerhead motif ribozyme molecules may specifically and efficiently catalyze endonucleolytic cleavage of sequences encoding HGPRP.

Specific ribozyme cleavage sites within any potential RNA target are initially identified by scanning the target molecule for ribozyme cleavage sites, including the following sequences: GUA, GUU, and GUC. Once identified, short RNA sequences of between 15 and 20 ribonucleotides, corresponding to the region of the target gene containing the cleavage site, may be evaluated for secondary structural features which may render the oligonucleotide inoperable. The suitability of candidate targets may also be evaluated by testing accessibility to hybridization with complementary oligonucleotides using ribonuclease protection assays.

Complementary ribonucleic acid molecules and ribozymes of the invention may be prepared by any method known in the art for the synthesis of nucleic acid molecules. These include techniques for chemically synthesizing oligonucleotides such as solid phase phosphoramidite chemical synthesis. Alternatively, RNA molecules may be generated by in vitro and in vivo transcription of DNA sequences encoding HGPRP. Such DNA sequences may be incorporated into a wide variety of vectors with suitable RNA polymerase promoters such as T7 or SP6. Alternatively, these cDNA constructs that synthesize complementary RNA, constitutively or inducibly, can be introduced into cell lines, cells, or tissues.

RNA molecules may be modified to increase intracellular stability and half-life. Possible

modifications include, but are not limited to. the addition of flanking sequences at the 5' and/or 3' ends of the molecule, or the use of phosphorothioate or 2' O-methyl rather than phosphodiesterase linkages within the backbone of the molecule. This concept is inherent in the production of PNAs and can be extended in all of these molecules by the inclusion of nontraditional bases such as inosine, queosine, and wybutosine, as well as acetyl-, methyl-, thio-, and similarly modified forms of adenine, cytidine, guanine, thymine, and uridine which are not as easily recognized by endogenous endonucleases.

Many methods for introducing vectors into cells or tissues are available and equally suitable for use in vivo, in vitro, and ex vivo. For ex vivo therapy, vectors may be introduced into stem cells taken from the patient and clonally propagated for autologous transplant back into that same patient. Delivery by transfection, by liposome injections, or by polycationic amino polymers may be achieved using methods which are well known in the art. (See, e.g., Goldman, C.K. et al. (1997) Nat. Biotechnol. 15:462-466.)

Any of the therapeutic methods described above may be applied to any subject in need of such therapy, including, for example, mammals such as dogs, cats, cows, horses, rabbits, monkeys, and most preferably, humans.

An additional embodiment of the invention relates to the administration of a pharmaceutical or sterile composition, in conjunction with a pharmaceutically acceptable carrier, for any of the therapeutic effects discussed above. Such pharmaceutical compositions may consist of HGPRP, antibodies to HGPRP, and mimetics, agonists, antagonists, or inhibitors of HGPRP. The compositions may be administered alone or in combination with at least one other agent, such as a stabilizing compound, which may be administered in any sterile, biocompatible pharmaceutical carrier including, but not limited to, saline, buffered saline, dextrose, and water. The compositions may be administered to a patient alone, or in combination with other agents, drugs, or hormones.

The pharmaceutical compositions utilized in this invention may be administered by any number of routes including, but not limited to, oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, intraventricular, transdermal, subcutaneous, intraperitoneal, intranasal, enteral, topical, sublingual, or rectal means.

In addition to the active ingredients, these pharmaceutical compositions may contain suitable pharmaceutically-acceptable carriers comprising excipients and auxiliaries which facilitate processing of the active compounds into preparations which can be used pharmaceutically. Further details on techniques for formulation and administration may be found in the latest edition of Remington's Pharmaceutical Sciences (Maack Publishing, Easton PA).

Pharmaceutical compositions for oral administration can be formulated using

pharmaceutically acceptable carriers well known in the art in dosages suitable for oral administration. Such carriers enable the pharmaceutical compositions to be formulated as tablets, pills, dragees, capsules, liquids, gels, syrups, slurries, suspensions, and the like, for ingestion by the patient.

Pharmaceutical preparations for oral use can be obtained through combining active compounds with solid excipient and processing the resultant mixture of granules (optionally, after grinding) to obtain tablets or dragee cores. Suitable auxiliaries can be added, if desired. Suitable excipients include carbohydrate or protein fillers, such as sugars, including lactose, sucrose, mannitol, and sorbitol; starch from corn, wheat, rice, potato, or other plants; cellulose, such as methyl cellulose, hydroxypropylmethyl-cellulose, or sodium carboxymethylcellulose; gums, including arabic and tragacanth; and proteins, such as gelatin and collagen. If desired, disintegrating or solubilizing agents may be added, such as the cross-linked polyvinyl pyrrolidone, agar, and alginic acid or a salt thereof, such as sodium alginate.

Dragee cores may be used in conjunction with suitable coatings, such as concentrated sugar solutions, which may also contain gum arabic, talc, polyvinylpyrrolidone, carbopol gel, polyethylene glycol, and/or titanium dioxide, lacquer solutions, and suitable organic solvents or solvent mixtures. Dyestuffs or pigments may be added to the tablets or dragee coatings for product identification or to characterize the quantity of active compound, i.e., dosage.

Pharmaceutical preparations which can be used orally include push-fit capsules made of gelatin, as well as soft, sealed capsules made of gelatin and a coating, such as glycerol or sorbitol. Push-fit capsules can contain active ingredients mixed with fillers or binders, such as lactose or starches, lubricants, such as talc or magnesium stearate, and, optionally, stabilizers. In soft capsules, the active compounds may be dissolved or suspended in suitable liquids, such as fatty oils, liquid, or liquid polyethylene glycol with or without stabilizers.

Pharmaceutical formulations suitable for parenteral administration may be formulated in aqueous solutions, preferably in physiologically compatible buffers such as Hanks' solution, Ringer's solution, or physiologically buffered saline. Aqueous injection suspensions may contain substances which increase the viscosity of the suspension, such as sodium carboxymethyl cellulose, sorbitol, or dextran. Additionally, suspensions of the active compounds may be prepared as appropriate oily injection suspensions. Suitable lipophilic solvents or vehicles include fatty oils, such as sesame oil, or synthetic fatty acid esters, such as ethyl oleate, triglycerides, or liposomes. Non-lipid polycationic amino polymers may also be used for delivery. Optionally, the suspension may also contain suitable stabilizers or agents to increase the solubility of the compounds and allow for the preparation of highly concentrated solutions.

For topical or nasal administration, penetrants appropriate to the particular barrier to be

permeated are used in the formulation. Such penetrants are generally known in the art.

The pharmaceutical compositions of the present invention may be manufactured in a manner that is known in the art, e.g., by means of conventional mixing, dissolving, granulating, dragee-making, levigating, emulsifying, encapsulating, entrapping, or lyophilizing processes.

The pharmaceutical composition may be provided as a salt and can be formed with many acids, including but not limited to, hydrochloric, sulfuric, acetic, lactic, tartaric, malic, and succinic acid. Salts tend to be more soluble in aqueous or other protonic solvents than are the corresponding free base forms. In other cases, the preferred preparation may be a lyophilized powder which may contain any or all of the following: 1 mM to 50 mM histidine, 0.1% to 2% sucrose, and 2% to 7% mannitol, at a pH range of 4.5 to 5.5, that is combined with buffer prior to use.

After pharmaceutical compositions have been prepared, they can be placed in an appropriate container and labeled for treatment of an indicated condition. For administration of HGPRP, such labeling would include amount, frequency, and method of administration.

Pharmaceutical compositions suitable for use in the invention include compositions wherein the active ingredients are contained in an effective amount to achieve the intended purpose. The determination of an effective dose is well within the capability of those skilled in the art.

For any compound, the therapeutically effective dose can be estimated initially either in cell culture assays, e.g., of neoplastic cells or in animal models such as mice, rats, rabbits, dogs, or pigs. An animal model may also be used to determine the appropriate concentration range and route of administration. Such information can then be used to determine useful doses and routes for administration in humans.

A therapeutically effective dose refers to that amount of active ingredient, for example HGPRP or fragments thereof, antibodies of HGPRP, and agonists, antagonists or inhibitors of HGPRP, which ameliorates the symptoms or condition. Therapeutic efficacy and toxicity may be determined by standard pharmaceutical procedures in cell cultures or with experimental animals, such as by calculating the ED_{50} (the dose therapeutically effective in 50% of the population) or LD_{50} (the dose lethal to 50% of the population) statistics. The dose ratio of toxic to therapeutic effects is the therapeutic index, and it can be expressed as the LD_{50}/ED_{50} ratio. Pharmaceutical compositions which exhibit large therapeutic indices are preferred. The data obtained from cell culture assays and animal studies are used to formulate a range of dosage for human use. The dosage contained in such compositions is preferably within a range of circulating concentrations that includes the ED_{50} with little or no toxicity. The dosage varies within this range depending upon the dosage form employed, the sensitivity of the patient, and the route of administration.

The exact dosage will be determined by the practitioner, in light of factors related to the subject requiring treatment. Dosage and administration are adjusted to provide sufficient levels of the active moiety or to maintain the desired effect. Factors which may be taken into account include the severity of the disease state, the general health of the subject, the age, weight, and gender of the subject, time and frequency of administration, drug combination(s), reaction sensitivities, and response to therapy. Long-acting pharmaceutical compositions may be administered every 3 to 4 days, every week, or biweekly depending on the half-life and clearance rate of the particular formulation.

Normal dosage amounts may vary from about 0.1 μg to 100,000 μg, up to a total dose of about 1 gram, depending upon the route of administration. Guidance as to particular dosages and methods of delivery is provided in the literature and generally available to practitioners in the art. Those skilled in the art will employ different formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations, etc.

DIAGNOSTICS

In another embodiment, antibodies which specifically bind HGPRP may be used for the diagnosis of disorders characterized by expression of HGPRP, or in assays to monitor patients being treated with HGPRP or agonists, antagonists, or inhibitors of HGPRP. Antibodies useful for diagnostic purposes may be prepared in the same manner as described above for therapeutics. Diagnostic assays for HGPRP include methods which utilize the antibody and a label to detect HGPRP in human body fluids or in extracts of cells or tissues. The antibodies may be used with or without modification, and may be labeled by covalent or non-covalent attachment of a reporter molecule. A wide variety of reporter molecules, several of which are described above, are known in the art and may be used.

A variety of protocols for measuring HGPRP, including ELISAs, RIAs, and FACS, are known in the art and provide a basis for diagnosing altered or abnormal levels of HGPRP expression. Normal or standard values for HGPRP expression are established by combining body fluids or cell extracts taken from normal mammalian subjects, preferably human, with antibody to HGPRP under conditions suitable for complex formation. The amount of standard complex formation may be quantitated by various methods, preferably by photometric means. Quantities of HGPRP expressed in subject, control, and disease samples from biopsied tissues are compared with the standard values. Deviation between standard and subject values establishes the parameters for diagnosing disease.

In another embodiment of the invention, the polynucleotides encoding HGPRP may be used for diagnostic purposes. The polynucleotides which may be used include oligonucleotide

sequences, complementary RNA and DNA molecules, and PNAs. The polynucleotides may be used to detect and quantitate gene expression in biopsied tissues in which expression of HGPRP may be correlated with disease. The diagnostic assay may be used to determine absence, presence, and excess expression of HGPRP, and to monitor regulation of HGPRP levels during therapeutic intervention.

In one aspect, hybridization with PCR probes which are capable of detecting polynucleotide sequences, including genomic sequences, encoding HGPRP or closely related molecules may be used to identify nucleic acid sequences which encode HGPRP. The specificity of the probe, whether it is made from a highly specific region, e.g., the 5' regulatory region, or from a less specific region, e.g., a conserved motif, and the stringency of the hybridization or amplification (maximal, high, intermediate, or low), will determine whether the probe identifies only naturally occurring sequences encoding HGPRP, allelic variants, or related sequences.

Probes may also be used for the detection of related sequences, and should preferably have at least 50% sequence identity to any of the HGPRP encoding sequences. The hybridization probes of the subject invention may be DNA or RNA and may be derived from the sequence of SEQ ID NO:7-12 or from genomic sequences including promoters, enhancers, and introns of the HGPRP gene.

Means for producing specific hybridization probes for DNAs encoding HGPRP include the cloning of polynucleotide sequences encoding HGPRP or HGPRP derivatives into vectors for the production of mRNA probes. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes <u>in vitro</u> by means of the addition of the appropriate RNA polymerases and the appropriate labeled nucleotides. Hybridization probes may be labeled by a variety of reporter groups, for example, by radionuclides such as ^{32}P or ^{35}S, or by enzymatic labels, such as alkaline phosphatase coupled to the probe via avidin/biotin coupling systems, and the like.

Polynucleotide sequences encoding HGPRP may be used for the diagnosis of disorders associated with expression of HGPRP. Examples of such disorders include, but are not limited to, a cell proliferative disorder such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia; cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus; an immune disorder such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult

respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis. dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis. Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; and a neurological disorder such as epilepsy, ischemic cerebrovascular disease, stroke, cerebral neoplasms. Alzheimer's disease, Pick's disease, Huntington's disease, dementia, Parkinson's disease and other extrapyramidal disorders, amyotrophic lateral sclerosis and other motor neuron disorders, progressive neural muscular atrophy, retinitis pigmentosa, hereditary ataxias, multiple sclerosis and other demyelinating diseases, bacterial and viral meningitis, brain abscess, subdural empyema, epidural abscess, suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral central nervous system disease; prion diseases including kuru, Creutzfeldt-Jakob disease, and Gerstmann-Straussler-Scheinker syndrome; fatal familial insomnia, nutritional and metabolic diseases of the nervous system, neurofibromatosis, tuberous sclerosis, cerebelloretinal hemangioblastomatosis, encephalotrigeminal syndrome, mental retardation and other developmental disorders of the central nervous system, cerebral palsy, neuroskeletal disorders, autonomic nervous system disorders, cranial nerve disorders, spinal cord diseases, muscular dystrophy and other neuromuscular disorders, peripheral nervous system disorders, dermatomyositis and polymyositis; inherited, metabolic, endocrine, and toxic myopathies; myasthenia gravis, periodic paralysis; mental disorders including mood, anxiety, and schizophrenic disorders; akathesia, amnesia, catatonia, diabetic neuropathy, tardive dyskinesia, dystonias, paranoid psychoses, postherpetic neuralgia, and Tourette's disorder. The polynucleotide sequences encoding HGPRP may be used in Southern or northern analysis, dot blot, or other membrane-based technologies; in PCR technologies; in dipstick, pin, and multiformat ELISA-like assays; and in microarrays utilizing fluids or tissues from patients to detect altered HGPRP expression. Such qualitative or quantitative methods are well known in the art.

In a particular aspect, the nucleotide sequences encoding HGPRP may be useful in assays that detect the presence of associated disorders, particularly those mentioned above. The

nucleotide sequences encoding HGPRP may be labeled by standard methods and added to a fluid or tissue sample from a patient under conditions suitable for the formation of hybridization complexes. After a suitable incubation period, the sample is washed and the signal is quantitated and compared with a standard value. If the amount of signal in the patient sample is significantly altered in comparison to a control sample then the presence of altered levels of nucleotide sequences encoding HGPRP in the sample indicates the presence of the associated disorder. Such assays may also be used to evaluate the efficacy of a particular therapeutic treatment regimen in animal studies, in clinical trials, or to monitor the treatment of an individual patient.

In order to provide a basis for the diagnosis of a disorder associated with expression of HGPRP, a normal or standard profile for expression is established. This may be accomplished by combining body fluids or cell extracts taken from normal subjects, either animal or human, with a sequence, or a fragment thereof, encoding HGPRP, under conditions suitable for hybridization or amplification. Standard hybridization may be quantified by comparing the values obtained from normal subjects with values from an experiment in which a known amount of a substantially purified polynucleotide is used. Standard values obtained in this manner may be compared with values obtained from samples from patients who are symptomatic for a disorder. Deviation from standard values is used to establish the presence of a disorder.

Once the presence of a disorder is established and a treatment protocol is initiated, hybridization assays may be repeated on a regular basis to determine if the level of expression in the patient begins to approximate that which is observed in the normal subject. The results obtained from successive assays may be used to show the efficacy of treatment over a period ranging from several days to months.

With respect to cancer, the presence of an abnormal amount of transcript (either under- or overexpressed) in biopsied tissue from an individual may indicate a predisposition for the development of the disease, or may provide a means for detecting the disease prior to the appearance of actual clinical symptoms. A more definitive diagnosis of this type may allow health professionals to employ preventative measures or aggressive treatment earlier thereby preventing the development or further progression of the cancer.

Additional diagnostic uses for oligonucleotides designed from the sequences encoding HGPRP may involve the use of PCR. These oligomers may be chemically synthesized, generated enzymatically, or produced in vitro. Oligomers will preferably contain a fragment of a polynucleotide encoding HGPRP, or a fragment of a polynucleotide complementary to the polynucleotide encoding HGPRP, and will be employed under optimized conditions for identification of a specific gene or condition. Oligomers may also be employed under less stringent conditions for detection or quantitation of closely related DNA or RNA sequences.

Methods which may also be used to quantitate the expression of HGPRP include radiolabeling or biotinylating nucleotides, coamplification of a control nucleic acid, and interpolating results from standard curves. (See, e.g., Melby, P.C. et al. (1993) J. Immunol. Methods 159:235-244; Duplaa, C. et al. (1993) Anal. Biochem. 212:229-236.) The speed of quantitation of multiple samples may be accelerated by running the assay in an ELISA format where the oligomer of interest is presented in various dilutions and a spectrophotometric or colorimetric response gives rapid quantitation.

In further embodiments, oligonucleotides or longer fragments derived from any of the polynucleotide sequences described herein may be used as targets in a microarray. The microarray can be used to monitor the expression level of large numbers of genes simultaneously and to identify genetic variants, mutations, and polymorphisms. This information may be used to determine gene function, to understand the genetic basis of a disorder, to diagnose a disorder, and to develop and monitor the activities of therapeutic agents.

Microarrays may be prepared, used, and analyzed using methods known in the art. (See, e.g., Brennan, T.M. et al. (1995) U.S. Patent No. 5,474,796; Schena, M. et al. (1996) Proc. Natl. Acad. Sci. USA 93:10614-10619; Baldeschweiler et al. (1995) PCT application WO95/251116; Shalon, D. et al. (1995) PCT application WO95/35505; Heller, R.A. et al. (1997) Proc. Natl. Acad. Sci. USA 94:2150-2155; and Heller, M.J. et al. (1997) U.S. Patent No. 5,605,662.)

In another embodiment of the invention, nucleic acid sequences encoding HGPRP may be used to generate hybridization probes useful in mapping the naturally occurring genomic sequence. The sequences may be mapped to a particular chromosome, to a specific region of a chromosome, or to artificial chromosome constructions, e.g., human artificial chromosomes (HACs), yeast artificial chromosomes (YACs), bacterial artificial chromosomes (BACs), bacterial P1 constructions, or single chromosome cDNA libraries. (See, e.g., Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355; Price, C.M. (1993) Blood Rev. 7:127-134; and Trask, B.J. (1991) Trends Genet. 7:149-154.)

Fluorescent in situ hybridization (FISH) may be correlated with other physical chromosome mapping techniques and genetic map data. (See, e.g., Heinz-Ulrich, et al. (1995) in Meyers, supra, pp. 965-968.) Examples of genetic map data can be found in various scientific journals or at the Online Mendelian Inheritance in Man (OMIM) site. Correlation between the location of the gene encoding HGPRP on a physical chromosomal map and a specific disorder, or a predisposition to a specific disorder, may help define the region of DNA associated with that disorder. The nucleotide sequences of the invention may be used to detect differences in gene sequences among normal, carrier, and affected individuals.

In situ hybridization of chromosomal preparations and physical mapping techniques, such

as linkage analysis using established chromosomal markers, may be used for extending genetic maps. Often the placement of a gene on the chromosome of another mammalian species, such as mouse, may reveal associated markers even if the number or arm of a particular human chromosome is not known. New sequences can be assigned to chromosomal arms by physical mapping. This provides valuable information to investigators searching for disease genes using positional cloning or other gene discovery techniques. Once the disease or syndrome has been crudely localized by genetic linkage to a particular genomic region, e.g., ataxia-telangiectasia to 11q22-23, any sequences mapping to that area may represent associated or regulatory genes for further investigation. (See, e.g., Gatti, R.A. et al. (1988) Nature 336:577-580.) The nucleotide sequence of the subject invention may also be used to detect differences in the chromosomal location due to translocation, inversion, etc., among normal, carrier, or affected individuals.

In another embodiment of the invention, HGPRP, its catalytic or immunogenic fragments, or oligopeptides thereof can be used for screening libraries of compounds in any of a variety of drug screening techniques. The fragment employed in such screening may be free in solution, affixed to a solid support, borne on a cell surface, or located intracellularly. The formation of binding complexes between HGPRP and the agent being tested may be measured.

Another technique for drug screening provides for high throughput screening of compounds having suitable binding affinity to the protein of interest. (See, e.g., Geysen, et al. (1984) PCT application WO84/03564.) In this method, large numbers of different small test compounds are synthesized on a solid substrate. The test compounds are reacted with HGPRP, or fragments thereof, and washed. Bound HGPRP is then detected by methods well known in the art. Purified HGPRP can also be coated directly onto plates for use in the aforementioned drug screening techniques. Alternatively, non-neutralizing antibodies can be used to capture the peptide and immobilize it on a solid support.

In another embodiment, one may use competitive drug screening assays in which neutralizing antibodies capable of binding HGPRP specifically compete with a test compound for binding HGPRP. In this manner, antibodies can be used to detect the presence of any peptide which shares one or more antigenic determinants with HGPRP.

In additional embodiments, the nucleotide sequences which encode HGPRP may be used in any molecular biology techniques that have yet to be developed, provided the new techniques rely on properties of nucleotide sequences that are currently known, including, but not limited to, such properties as the triplet genetic code and specific base pair interactions.

Without further elaboration, it is believed that one skilled in the art can, using the preceding description, utilize the present invention to its fullest extent. The following preferred

specific embodiments are, therefore, to be construed as merely illustrative, and not limitative of the remainder of the disclosure in any way whatsoever.

The disclosures of all patents, applications, and publications mentioned above and below, in particular U.S. Ser. No. 09/156,513, are hereby expressly incorporated by reference.

EXAMPLES

I. Construction of cDNA Libraries

RNA was purchased from Clontech or isolated from tissues described in Table 4. Some tissues were homogenized and lysed in guanidinium isothiocyanate, while others were homogenized and lysed in phenol or in a suitable mixture of denaturants, such as TRIZOL (Life Technologies), a monophasic solution of phenol and guanidine isothiocyanate. The resulting lysates were centrifuged over CsCl cushions or extracted with chloroform. RNA was precipitated from the lysates with either isopropanol or sodium acetate and ethanol, or by other routine methods.

Phenol extraction and precipitation of RNA were repeated as necessary to increase RNA purity. In some cases, RNA was treated with DNase. For most libraries, poly(A+) RNA was isolated using oligo d(T)-coupled paramagnetic particles (Promega), OLIGOTEX latex particles (Qiagen, Valencia CA), or an OLIGOTEX mRNA purification kit (Qiagen). Alternatively, RNA was isolated directly from tissue lysates using other RNA isolation kits, e.g., the POLY(A)PURE mRNA purification kit (Ambion, Austin TX).

In some cases, Stratagene was provided with RNA and constructed the corresponding cDNA libraries. Otherwise, cDNA was synthesized and cDNA libraries were constructed with the UNIZAP vector system (Stratagene) or SUPERSCRIPT plasmid system (Life Technologies), using the recommended procedures or similar methods known in the art. (See, e.g., Ausubel, 1997, supra, units 5.1-6.6.) Reverse transcription was initiated using oligo d(T) or random primers. Synthetic oligonucleotide adapters were ligated to double stranded cDNA, and the cDNA was digested with the appropriate restriction enzyme or enzymes. For most libraries, the cDNA was size-selected (300-1000 bp) using SEPHACRYL S1000, SEPHAROSE CL2B, or SEPHAROSE CL4B column chromatography (Amersham Pharmacia Biotech) or preparative agarose gel electrophoresis. cDNAs were ligated into compatible restriction enzyme sites of the polylinker of a suitable plasmid, e.g., PBLUESCRIPT plasmid (Stratagene), PSPORT1 plasmid (Life Technologies), or pINCY (Incyte Pharmaceuticals, Palo Alto CA). Recombinant plasmids were transformed into competent E. coli cells including XL1-Blue, XL1-BlueMRF, or SOLR from Stratagene, or DH5α, DH10B, or ElectroMAX DH10B from Life Technologies.

II. Isolation of cDNA Clones

Plasmids were recovered from host cells by in vivo excision, using the UNIZAP vector system (Stratagene) or cell lysis. Plasmids were purified using at least one of the following: a MAGIC or WIZARD MINIPREPS DNA purification system (Promega); an AGTC MINIPREP purification kit (Edge Biosystems, Gaithersburg MD); and QIAWELL 8 Plasmid, QIAWELL 8 Plus Plasmid, QIAWELL 8 Ultra Plasmid purification systems or the REAL Prep 96 plasmid kit from Qiagen. Following precipitation, plasmids were resuspended in 0.1 ml of distilled water and stored, with or without lyophilization, at 4°C.

Alternatively, plasmid DNA was amplified from host cell lysates using direct link PCR in a high-throughput format (Rao, V.B. (1994) Anal. Biochem. 216:1-14). Host cell lysis and thermal cycling steps were carried out in a single reaction mixture. Samples were processed and stored in 384-well plates, and the concentration of amplified plasmid DNA was quantified fluorometrically using PICOGREEN dye (Molecular Probes, Eugene OR) and a Fluoroskan II fluorescence scanner (Labsystems Oy, Helsinki, Finland).

III. Sequencing and Analysis

The cDNAs were prepared for sequencing using the ABI CATALYST 800 (PE Biosystems) or the HYDRA microdispenser (Robbins Scientific) or MICROLAB 2200 (Hamilton) systems in combination with the DNA ENGINE thermal cyclers (MJ Research). The cDNAs were sequenced using the ABI PRISM 373 or 377 sequencing systems (PE Biosystems) and standard ABI protocols, base calling software, and kits. In one alternative, cDNAs were sequenced using the MEGABACE 1000 DNA sequencing system (Amersham Pharmacia Biotech). In another alternative, the cDNAs were amplified and sequenced using the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (PE Biosystems). In yet another alternative, cDNAs were sequenced using solutions and dyes from Amersham Pharmacia Biotech. Reading frames for the ESTs were determined using standard methods (reviewed in Ausubel, 1997, supra, unit 7.7). Some of the cDNA sequences were selected for extension using the techniques disclosed in Example V.

The polynucleotide sequences derived from cDNA, extension, and shotgun sequencing were assembled and analyzed using a combination of software programs which utilize algorithms well known to those skilled in the art. Table 5 summarizes the software programs, descriptions, references, and threshold parameters used. The first column of Table 5 shows the tools, programs, and algorithms used, the second column provides a brief description thereof, the third column presents the references which are incorporated by reference herein, and the fourth column presents, where applicable, the scores, probability values, and other parameters used to evaluate the strength of a match between two sequences (the higher the probability the greater the

homology). Sequences were analyzed using MACDNASIS PRO software (Hitachi Software Engineering, S. San Francisco CA) and LASERGENE software (DNASTAR).

The polynucleotide sequences were validated by removing vector, linker, and polyA sequences and by masking ambiguous bases, using algorithms and programs based on BLAST, dynamic programing, and dinucleotide nearest neighbor analysis. The sequences were then queried against a selection of public databases such as GENBANK primate, rodent, mammalian, vertebrate, and eukaryote databases, and BLOCKS to acquire annotation, using programs based on BLAST, FASTA, and BLIMPS. The sequences were assembled into full length polynucleotide sequences using programs based on Phred, Phrap, and Consed, and were screened for open reading frames using programs based on GeneMark, BLAST, and FASTA. The full length polynucleotide sequences were translated to derive the corresponding full length amino acid sequences, and these full length sequences were subsequently analyzed by querying against databases such as the GENBANK databases (described above), SWISSPROT, BLOCKS, PRINTS, PFAM, and PROSITE.

The programs described above for the assembly and analysis of full length polynucleotide and amino acid sequences were also used to identify polynucleotide sequence fragments from SEQ ID NO:7-12. Fragments from about 20 to about 4000 nucleotides which are useful in hybridization and amplification technologies were described in The Invention section above.

IV. Northern Analysis

Northern analysis is a laboratory technique used to detect the presence of a transcript of a gene and involves the hybridization of a labeled nucleotide sequence to a membrane on which RNAs from a particular cell type or tissue have been bound. (See, e.g., Sambrook, supra, ch. 7; Ausubel, 1995, supra, ch. 4 and 16.)

Analogous computer techniques applying BLAST were used to search for identical or related molecules in nucleotide databases such as GENBANK or LIFESEQ database (Incyte Pharmaceuticals). This analysis is much faster than multiple membrane-based hybridizations. In addition, the sensitivity of the computer search can be modified to determine whether any particular match is categorized as exact or similar. The basis of the search is the product score, which is defined as:

$$\frac{\%\ \text{sequence identity} \times \%\ \text{maximum BLAST score}}{100}$$

The product score takes into account both the degree of similarity between two sequences and the length of the sequence match. For example, with a product score of 40, the match will be exact within a 1% to 2% error, and, with a product score of 70, the match will be exact. Similar molecules are usually identified by selecting those which show product scores between 15 and 40,

although lower scores may identify related molecules.

The results of northern analyses are reported a percentage distribution of libraries in which the transcript encoding HGPRP occurred. Analysis involved the categorization of cDNA libraries by organ/tissue and disease. The organ/tissue categories included cardiovascular, dermatologic, developmental, endocrine, gastrointestinal, hematopoietic/immune, musculoskeletal, nervous, reproductive, and urologic. The disease categories included cancer, inflammation/trauma, fetal, neurological, and pooled. For each category, the number of libraries expressing the sequence of interest was counted and divided by the total number of libraries across all categories. Percentage values of tissue-specific and disease expression are reported in Table 3.

V. Extension of HGPRP Encoding Polynucleotides

The full length nucleic acid sequence of SEQ ID NO:7-12 was produced by extension of an appropriate fragment of the full length molecule using oligonucleotide primers designed from this fragment. One primer was synthesized to initiate 5' extension of the known fragment, and the other primer, to initiate 3' extension of the known fragment. The initial primers were designed using OLIGO 4.06 software (National Biosciences), or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the target sequence at temperatures of about 68°C to about 72°C. Any stretch of nucleotides which would result in hairpin structures and primer-primer dimerizations was avoided.

Selected human cDNA libraries were used to extend the sequence. If more than one extension was necessary or desired, additional or nested sets of primers were designed.

High fidelity amplification was obtained by PCR using methods well known in the art. PCR was performed in 96-well plates using the DNA ENGINE thermal cycler (MJ Research). The reaction mix contained DNA template, 200 nmol of each primer, reaction buffer containing Mg^{2+}, $(NH_4)_2SO_4$, and β-mercaptoethanol, Taq DNA polymerase (Amersham Pharmacia Biotech), ELONGASE enzyme (Life Technologies), and Pfu DNA polymerase (Stratagene), with the following parameters for primer pair PCI A and PCI B: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C. In the alternative, the parameters for primer pair T7 and SK+ were as follows: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 57°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C.

The concentration of DNA in each well was determined by dispensing 100 µl PICOGREEN quantitation reagent (0.25% (v/v) PICOGREEN; Molecular Probes, Eugene OR) dissolved in 1X TE and 0.5 µl of undiluted PCR product into each well of an opaque fluorimeter plate (Corning Costar, Acton MA), allowing the DNA to bind to the reagent. The plate was scanned in a Fluoroskan II (Labsystems Oy, Helsinki, Finland) to measure the fluorescence of the

sample and to quantify the concentration of DNA. A 5 μl to 10 μl aliquot of the reaction mixture was analyzed by electrophoresis on a 1 % agarose mini-gel to determine which reactions were successful in extending the sequence.

The extended nucleotides were desalted and concentrated, transferred to 384-well plates, digested with CviJI cholera virus endonuclease (Molecular Biology Research, Madison WI), and sonicated or sheared prior to religation into pUC 18 vector (Amersham Pharmacia Biotech). For shotgun sequencing, the digested nucleotides were separated on low concentration (0.6 to 0.8%) agarose gels, fragments were excised, and agar digested with Agar ACE (Promega). Extended clones were religated using T4 ligase (New England Biolabs, Beverly MA) into pUC 18 vector (Amersham Pharmacia Biotech), treated with Pfu DNA polymerase (Stratagene) to fill-in restriction site overhangs, and transfected into competent E. coli cells. Transformed cells were selected on antibiotic-containing media, individual colonies were picked and cultured overnight at 37°C in 384-well plates in LB/2x carb liquid media.

The cells were lysed, and DNA was amplified by PCR using Taq DNA polymerase (Amersham Pharmacia Biotech) and Pfu DNA polymerase (Stratagene) with the following parameters: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 72°C, 2 min; Step 5: steps 2, 3, and 4 repeated 29 times; Step 6: 72°C, 5 min; Step 7: storage at 4°C. DNA was quantified by PICOGREEN reagent (Molecular Probes) as described above. Samples with low DNA recoveries were reamplified using the same conditions as described above. Samples were diluted with 20% dimethysulphoxide (1:2, v/v), and sequenced using DYENAMIC energy transfer sequencing primers and the DYENAMIC DIRECT kit (Amersham Pharmacia Biotech) or the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (PE Biosystems).

In like manner, the nucleotide sequence of SEQ ID NO:7-12 is used to obtain 5' regulatory sequences using the procedure above, oligonucleotides designed for such extension, and an appropriate genomic library.

VI. Labeling and Use of Individual Hybridization Probes

Hybridization probes derived from SEQ ID NO:7-12 are employed to screen cDNAs, genomic DNAs, or mRNAs. Although the labeling of oligonucleotides, consisting of about 20 base pairs, is specifically described, essentially the same procedure is used with larger nucleotide fragments. Oligonucleotides are designed using state-of-the-art software such as OLIGO 4.06 software (National Biosciences) and labeled by combining 50 pmol of each oligomer, 250 μCi of [γ-^{32}P] adenosine triphosphate (Amersham Pharmacia Biotech), and T4 polynucleotide kinase (DuPont NEN, Boston MA). The labeled oligonucleotides are substantially purified using a SEPHADEX G-25 superfine size exclusion dextran bead column (Amersham Pharmacia Biotech). An aliquot containing 10^7 counts per minute of the labeled probe is used in a typical membrane-

based hybridization analysis of human genomic DNA digested with one of the following endonucleases: Ase I, Bgl II, Eco RI, Pst I, Xba I, or Pvu II (DuPont NEN).

The DNA from each digest is fractionated on a 0.7% agarose gel and transferred to nylon membranes (NYTRAN PLUS, Schleicher & Schuell, Durham NH). Hybridization is carried out for 16 hours at 40°C. To remove nonspecific signals, blots are sequentially washed at room temperature under increasingly stringent conditions up to 0.1 x saline sodium citrate and 0.5% sodium dodecyl sulfate. After XOMAT-AR film (Eastman Kodak, Rochester NY) is exposed to the blots for several hours, hybridization patterns are compared.

VII. Microarrays

A chemical coupling procedure and an ink jet device can be used to synthesize array elements on the surface of a substrate. (See, e.g., Baldeschweiler, supra.) An array analogous to a dot or slot blot may also be used to arrange and link elements to the surface of a substrate using thermal, UV, chemical, or mechanical bonding procedures. A typical array may be produced by hand or using available methods and machines and contain any appropriate number of elements. After hybridization, nonhybridized probes are removed and a scanner used to determine the levels and patterns of fluorescence. The degree of complementarity and the relative abundance of each probe which hybridizes to an element on the microarray may be assessed through analysis of the scanned images.

Full-length cDNAs, Expressed Sequence Tags (ESTs), or fragments thereof may comprise the elements of the microarray. Fragments suitable for hybridization can be selected using software well known in the art such as LASERGENE software (DNASTAR). Full-length cDNAs, ESTs, or fragments thereof corresponding to one of the nucleotide sequences of the present invention, or selected at random from a cDNA library relevant to the present invention, are arranged on an appropriate substrate, e.g., a glass slide. The cDNA is fixed to the slide using, e.g., UV cross-linking followed by thermal and chemical treatments and subsequent drying. (See, e.g., Schena, M. et al. (1995) Science 270:467-470; Shalon, D. et al. (1996) Genome Res. 6:639-645.) Fluorescent probes are prepared and used for hybridization to the elements on the substrate. The substrate is analyzed by procedures described above.

VIII. Complementary Polynucleotides

Sequences complementary to the HGPRP-encoding sequences, or any parts thereof, are used to detect, decrease, or inhibit expression of naturally occurring HGPRP. Although use of oligonucleotides comprising from about 15 to 30 base pairs is described, essentially the same procedure is used with smaller or with larger sequence fragments. Appropriate oligonucleotides are designed using OLIGO 4.06 software (National Biosciences) and the coding sequence of HGPRP. To inhibit transcription, a complementary oligonucleotide is designed from the most

unique 5' sequence and used to prevent promoter binding to the coding sequence. To inhibit translation, a complementary oligonucleotide is designed to prevent ribosomal binding to the HGPRP-encoding transcript.

IX. Expression of HGPRP

Expression and purification of HGPRP is achieved using bacterial or virus-based expression systems. For expression of HGPRP in bacteria, cDNA is subcloned into an appropriate vector containing an antibiotic resistance gene and an inducible promoter that directs high levels of cDNA transcription. Examples of such promoters include, but are not limited to, the *trp-lac* (*tac*) hybrid promoter and the T5 or T7 bacteriophage promoter in conjunction with the *lac* operator regulatory element. Recombinant vectors are transformed into suitable bacterial hosts, e.g., BL21(DE3). Antibiotic resistant bacteria express HGPRP upon induction with isopropyl beta-D-thiogalactopyranoside (IPTG). Expression of HGPRP in eukaryotic cells is achieved by infecting insect or mammalian cell lines with recombinant Autographica californica nuclear polyhedrosis virus (AcMNPV), commonly known as baculovirus. The nonessential polyhedrin gene of baculovirus is replaced with cDNA encoding HGPRP by either homologous recombination or bacterial-mediated transposition involving transfer plasmid intermediates. Viral infectivity is maintained and the strong polyhedrin promoter drives high levels of cDNA transcription. Recombinant baculovirus is used to infect Spodoptera frugiperda (Sf9) insect cells in most cases, or human hepatocytes, in some cases. Infection of the latter requires additional genetic modifications to baculovirus. (See Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945.)

In most expression systems, HGPRP is synthesized as a fusion protein with, e.g., glutathione S-transferase (GST) or a peptide epitope tag, such as FLAG or 6-His, permitting rapid, single-step, affinity-based purification of recombinant fusion protein from crude cell lysates. GST, a 26-kilodalton enzyme from Schistosoma japonicum, enables the purification of fusion proteins on immobilized glutathione under conditions that maintain protein activity and antigenicity (Amersham Pharmacia Biotech). Following purification, the GST moiety can be proteolytically cleaved from HGPRP at specifically engineered sites. FLAG, an 8-amino acid peptide, enables immunoaffinity purification using commercially available monoclonal and polyclonal anti-FLAG antibodies (Eastman Kodak). 6-His, a stretch of six consecutive histidine residues, enables purification on metal-chelate resins (QIAGEN). Methods for protein expression and purification are discussed in Ausubel (1995, supra, ch 10 and 16). Purified HGPRP obtained by these methods can be used directly in the following activity assay.

X. Demonstration of HGPRP Activity

GPCR activity of HGPRP is determined in a ligand-binding assay using candidate ligand

molecules in the presence of ^{125}I-labeled HGPRP. HGPRP is labeled with ^{125}I Bolton-Hunter reagent. (See, e.g., Bolton, A.E. and W.M. Hunter (1973) Biochem. J. 133:529-539.) Candidate ligand molecules previously arrayed in the wells of a multi-well plate are incubated with the labeled HGPRP, washed, and any wells with labeled HGPRP complex are assayed. Data obtained using different concentrations of HGPRP are used to calculate values for the number, affinity, and association of HGPRP with the ligand molecules.

XI. Functional Assays

HGPRP function is assessed by expressing the sequences encoding HGPRP at physiologically elevated levels in mammalian cell culture systems. cDNA is subcloned into a mammalian expression vector containing a strong promoter that drives high levels of cDNA expression. Vectors of choice include pCMV SPORT plasmid (Life Technologies) and pCR3.1 plasmid (Invitrogen, Carlsbad CA), both of which contain the cytomegalovirus promoter. 5-10 μg of recombinant vector are transiently transfected into a human cell line, preferably of endothelial or hematopoietic origin, using either liposome formulations or electroporation. 1-2 μg of an additional plasmid containing sequences encoding a marker protein are co-transfected. Expression of a marker protein provides a means to distinguish transfected cells from nontransfected cells and is a reliable predictor of cDNA expression from the recombinant vector. Marker proteins of choice include, e.g., Green Fluorescent Protein (GFP; Clontech), CD64, or a CD64-GFP fusion protein. Flow cytometry (FCM), an automated, laser optics-based technique, is used to identify transfected cells expressing GFP or CD64-GFP, and to evaluate properties, for example, their apoptotic state. FCM detects and quantifies the uptake of fluorescent molecules that diagnose events preceding or coincident with cell death. These events include changes in nuclear DNA content as measured by staining of DNA with propidium iodide; changes in cell size and granularity as measured by forward light scatter and 90 degree side light scatter; down-regulation of DNA synthesis as measured by decrease in bromodeoxyuridine uptake; alterations in expression of cell surface and intracellular proteins as measured by reactivity with specific antibodies; and alterations in plasma membrane composition as measured by the binding of fluorescein-conjugated Annexin V protein to the cell surface. Methods in flow cytometry are discussed in Ormerod, M.G. (1994) Flow Cytometry, Oxford, New York NY.

The influence of HGPRP on gene expression can be assessed using highly purified populations of cells transfected with sequences encoding HGPRP and either CD64 or CD64-GFP. CD64 and CD64-GFP are expressed on the surface of transfected cells and bind to conserved regions of human immunoglobulin G (IgG). Transfected cells are efficiently separated from nontransfected cells using magnetic beads coated with either human IgG or antibody against CD64 (DYNAL, Lake Success NY). mRNA can be purified from the cells using methods well known

by those of skill in the art. Expression of mRNA encoding HGPRP and other genes of interest can be analyzed by northern analysis or microarray techniques.

XII. Production of HGPRP Specific Antibodies

HGPRP substantially purified using polyacrylamide gel electrophoresis (PAGE; see. e.g., Harrington. M.G. (1990) Methods Enzymol. 182:488-495), or other purification techniques, is used to immunize rabbits and to produce antibodies using standard protocols.

Alternatively, the HGPRP amino acid sequence is analyzed using LASERGENE software (DNASTAR) to determine regions of high immunogenicity, and a corresponding oligopeptide is synthesized and used to raise antibodies by means known to those of skill in the art. Methods for selection of appropriate epitopes, such as those near the C-terminus or in hydrophilic regions are well described in the art. (See, e.g., Ausubel, 1995, supra, ch. 11.)

Typically, oligopeptides 15 residues in length are synthesized using an ABI 431A peptide synthesizer (PE Biosystems) using fmoc-chemistry and coupled to KLH (Sigma-Aldrich, St. Louis MO) by reaction with N-maleimidobenzoyl-N-hydroxysuccinimide ester (MBS) to increase immunogenicity. (See, e.g., Ausubel, 1995, supra.) Rabbits are immunized with the oligopeptide-KLH complex in complete Freund's adjuvant. Resulting antisera are tested for antipeptide activity by, for example, binding the peptide to plastic, blocking with 1% BSA, reacting with rabbit antisera, washing, and reacting with radio-iodinated goat anti-rabbit IgG.

XIII. Purification of Naturally Occurring HGPRP Using Specific Antibodies

Naturally occurring or recombinant HGPRP is substantially purified by immunoaffinity chromatography using antibodies specific for HGPRP. An immunoaffinity column is constructed by covalently coupling anti-HGPRP antibody to an activated chromatographic resin, such as CNBr-activated SEPHAROSE (Amersham Pharmacia Biotech). After the coupling, the resin is blocked and washed according to the manufacturer's instructions.

Media containing HGPRP are passed over the immunoaffinity column, and the column is washed under conditions that allow the preferential absorbance of HGPRP (e.g., high ionic strength buffers in the presence of detergent). The column is eluted under conditions that disrupt antibody/HGPRP binding (e.g., a buffer of pH 2 to pH 3, or a high concentration of a chaotrope, such as urea or thiocyanate ion), and HGPRP is collected.

XIV. Identification of Molecules Which Interact with HGPRP

HGPRP, or biologically active fragments thereof, are labeled with ^{125}I Bolton-Hunter reagent. (See, e.g., Bolton and Hunter, supra.) Candidate molecules previously arrayed in the wells of a multi-well plate are incubated with the labeled HGPRP, washed, and any wells with labeled HGPRP complex are assayed. Data obtained using different concentrations of HGPRP are used to calculate values for the number. affinity, and association of HGPRP with the candidate

molecules.

Various modifications and variations of the described methods and systems of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with specific preferred embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the described modes for carrying out the invention which are obvious to those skilled in molecular biology or related fields are intended to be within the scope of the following claims.

Table 1

Polypeptide SEQ ID NO:	Nucleotide SEQ ID NO:	Clone ID	Library	Fragments
1	7	1258981	MENITUT03	1258981H1 (MENITUT03), 1442823R1 (THYRNOT03), 1962119T6 (BRSTNOT04), 2059242R6 (OVARNOT03), SATA01180F1, SARB01556F1, SARA01967F1
2	8	1459432	COLNFET02	1459432H1 (COLNFET02), 1459432R1 (COLNFET02), 1459432X12 (COLNFET02), 3001554F6 (TLYMNOT06), SAAC00257R1, SAAB00250R1, SAAB00523R1
3	9	2214673	SINTFET03	1482004T1 (CORPNOT02), 2214673H1 (SINTFET03), 3073644H1 (BONEUNT01), 3573501F6 (BRONNOT01), 4618526H1 (BRAYDIT01), 4857037H1 (BRSTTUT22), 5025086H1 (OVARNON03)
4	10	2488822	KIDNTUT13	153210R6 (THP1PLB02), 2488822H1 (KIDNTUT13), 2488822X308B1 (KIDNTUT13), 2488822X310D1 (KIDNTUT13), 3558664T6 (LUNGNOT31)
5	11	2705201	PONSAZT01	2705201H1 (PONSAZT01), 2705201X325F1 (PONSAZT01), 1262948X325F1 (SYNORAT05), 3141184H1 (SMCCNOT02), 384797R6 (HYPONOB01)
6	12	3036563	PENCNOT02	3036563H1 (PENCNOT02), 4457161H1 (HEAADIR01), SZAH00352F1, SZAH02656F1, SZAH01730F1, SZAH03622F1, SZAH01163F1, SZAH02669F1, SZAH00249F1

Table 2

SEQ ID NO:	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences	Identification	Analytical Methods
1	441	S85 T164 T274 S306 S344 T81 S118 T407 Y312 Y387	N191 N405	M1-A23, I51-V72, C116-A145, I156-L175, M207-P229, G242-T264, E330-K341	Metabotropic glutamate GPCR	BLOCKS, HMM, MOTIFS, PRINTS, SPSCAN
2	353	S158 T255 S86 T120 S151 S243 S246 T251 T317 S325	N13 N16 N23 N58 N84	I42-V66, P78-M99, W109-I149, V159-L180, T209-L232, V254-T278, Y293-R319	Somatostatin receptor-like GPCR	BLAST, BLOCKS, HMM, MOTIFS, PFAM, PRINTS, PROFILESCAN
3	333	T60 T218 S89 S172 T224	N8 N110 N300	Y44-L74, P62-H83, F109-R131, N143-L164, A231-G255, K278-P304	Rhodopsin-like GPCR	BLAST, BLOCKS, HMM, MOTIFS, PFAM, PRINTS
4	396	S36 S187 T251 S27 T323 S389	N7	I46-P70, Y79-I100, L117-F157, R166-S187, S219-F242, L265-L289, S302-K328	Rhodopsin-like GPCR	BLAST, BLOCKS, HMM, MOTIFS, PFAM, PRINTS, PROFILESCAN
5	403	S360 S368 S47 T318 S337 S5 T33 S123 T398	N30 N352	I57-L78, G94-E117, C122-V151, L162-L181, M198-F220, G233-L255	Metabotropic glutamate GPCR	BLOCKS, HMM, MOTIFS, PRINTS
6	807	T129 S155 S172 S201 S322 S347 S409 S662 S787 S794 S117 T166 T271 T402 T583 T587 T618 S771	N88 N110 N127 N281 N392 N424 N443 N505 N647 N785 N798	N425-T452, I475-W499, A549-L572, F636-N647, Q677-G696, H709-W730	Secretin-like GPCR	BLAST, BLOCKS, HMM, MOTIFS, PRINTS

Table 3

Nucleotide SEQ ID NO:	Tissue Expression (Fraction of Total)	Disease or Condition (Fraction of Total)	Vector
7	Reproductive (0.306) ardiovascular (0.177) Gastrointestinal (0.177)	Cell Proliferation (0.839) Inflammation (0.162)	pINCY
8	Reproductive (0.333) Urologic (0.222) Developmental (0.111)	Cell Proliferation (0.667) Inflammation (0.222)	pINCY
9	Nervous (0.286) Cardiovascular (0.190) Hematopoietic/Immune (0.190)	Cell Proliferation (0.428) Inflammation (0.333) Neurological (0.143)	pINCY
10	Developmental (0.333) Cardiovascular (0.167) Hematopoietic/Immune (0.167)	Cell Proliferation (0.833) Inflammation (0.334)	pINCY
11	Nervous (0.632) Reproductive (0.158) Musculoskeletal (0.105)	Cell Proliferation (0.368) Inflammation (0.316)	pINCY
12	Cardiovascular (0.244) Reproductive (0.244) Urologic (0.122)	Cell Proliferation (0.732) Inflammation (0.341)	pINCY

Table 4

Nucleotide SEQ ID NO:	Clone ID	Library	Library Comment
7	1258981	MENITUT03	Library was constructed using RNA isolated from brain meningioma tissue removed from a 35-year-old Caucasian female during excision of a cerebral meningeal lesion. Pathology indicated a benign neoplasm in the right cerebellopontine angle of the brain. Patient history included hypothyroidism. Family history included myocardial infarction and breast cancer.
8	1459432	COLNFET02	Library was constructed using RNA isolated from the colon tissue of a Caucasian female fetus, who died at 20 weeks' gestation.
9	2214673	SINTFET03	Library was constructed using RNA isolated from small intestine tissue removed from a Caucasian female fetus who died at 20 weeks' gestation.
10	2488822	KIDNTUT13	Library was constructed using RNA isolated from kidney tumor tissue removed from a 51-year-old Caucasian female during a nephroureterectomy. Pathology indicated a grade 3 renal cell carcinoma. Patient history included depressive disorder, hypoglycemia, and uterine endometriosis. Family history included calculus of the kidney, colon cancer, and type II diabetes.
11	2705201	PONSAZT01	Library was constructed using RNA isolated from diseased pons tissue removed from the brain of a 74-year-old Caucasian male who died from Alzheimer's disease.
12	3036563	PENCNOT02	Library was constructed using RNA isolated from penis right corpus cavernosum tissue removed from a male.

Table 5

Program	Description	Reference	Parameter Threshold
ABI FACTURA	A program that removes vector sequences and masks ambiguous bases in nucleic acid sequences.	Perkin-Elmer Applied Biosystems, Foster City, CA.	
ABI/PARACEL FDF	A Fast Data Finder useful in comparing and annotating amino acid or nucleic acid sequences.	Perkin-Elmer Applied Biosystems, Foster City, CA; Paracel Inc., Pasadena, CA.	Mismatch <50%
ABI AutoAssembler	A program that assembles nucleic acid sequences.	Perkin-Elmer Applied Biosystems, Foster City, CA.	
BLAST	A Basic Local Alignment Search Tool useful in sequence similarity search for amino acid and nucleic acid sequences. BLAST includes five functions: blastp, blastn, blastx, tblastn, and tblastx.	Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410; Altschul, S.F. et al. (1997) Nucleic Acids Res. 25: 3389-3402.	*ESTs:* Probability value= 1.0E-8 or less *Full Length sequences:* Probability value= 1.0E-10 or less
FASTA	A Pearson and Lipman algorithm that searches for similarity between a query sequence and a group of sequences of the same type. FASTA comprises as least five functions: fasta, tfasta, fastx, tfastx, and ssearch.	Pearson, W.R. and D.J. Lipman (1988) Proc. Natl. Acad Sci. 85:2444-2448; Pearson, W.R. (1990) Methods Enzymol. 183: 63-98; and Smith, T.F. and M. S. Waterman (1981) Adv. Appl. Math. 2:482-489.	*ESTs:* fasta E value=1.06E-6 *Assembled ESTs:* fasta Identity= 95% or greater and Match length=200 bases or greater; fastx E value=1.0E-8 or less *Full Length sequences:* fastx score=100 or greater
BLIMPS	A BLocks IMProved Searcher that matches a sequence against those in BLOCKS and PRINTS databases to search for gene families, sequence homology, and structural fingerprint regions.	Henikoff, S and J.G. Henikoff, Nucl. Acid Res., 19:6565-72, 1991. J.G. Henikoff and S. Henikoff (1996) Methods Enzymol. 266:88-105; and Attwood, T.K. et al. (1997) J. Chem. Inf. Comput. Sci. 37: 417-424.	Score=1000 or greater; Ratio of Score/Strength = 0.75 or larger; and Probability value= 1.0E-3 or less
PFAM	A Hidden Markov Models-based application useful for protein family search.	Krogh, A. et al. (1994) J. Mol. Biol., 235:1501-1531; Sonnhammer, E.L.L. et al. (1988) Nucleic Acids Res. 26:320-322.	Score=10-50 bits, depending on individual protein families

Table 5 cont.

Program	Description	Reference	Parameter Threshold
ProfileScan	An algorithm that searches for structural and sequence motifs in protein sequences that match sequence patterns defined in Prosite.	Gribskov, M. et al. (1988) CABIOS 4:61-66; Gribskov, et al. (1989) Methods Enzymol. 183:146-159; Bairoch, A. et al. (1997) Nucleic Acids Res. 25: 217-221.	Score= 4.0 or greater
Phred	A base-calling algorithm that examines automated sequencer traces with high sensitivity and probability.	Ewing, B. et al. (1998) Genome Res. 8:175-185; Ewing, B. and P. Green (1998) Genome Res. 8:186-194.	
Phrap	A Phils Revised Assembly Program including SWAT and CrossMatch, programs based on efficient implementation of the Smith-Waterman algorithm, useful in searching sequence homology and assembling DNA sequences.	Smith, T.F. and M. S. Waterman (1981) Adv. Appl. Math. 2:482-489; Smith, T.F. and M. S. Waterman (1981) J. Mol. Biol. 147:195-197; and Green, P., University of Washington, Seattle, WA.	Score= 120 or greater; Match length= 56 or greater
Consed	A graphical tool for viewing and editing Phrap assemblies	Gordon, D. et al. (1998) Genome Res. 8:195-202.	
SPScan	A weight matrix analysis program that scans protein sequences for the presence of secretory signal peptides.	Nielson, H. et al. (1997) Protein Engineering 10:1-6; Claverie, J.M. and S. Audic (1997) CABIOS 12: 431-439.	Score=5 or greater
Motifs	A program that searches amino acid sequences for patterns that matched those defined in Prosite.	Bairoch et al. supra; Wisconsin Package Program Manual, version 9, page M51-59, Genetics Computer Group, Madison, WI.	

What is claimed is:

1. A substantially purified polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, and SEQ ID NO:6, and fragments thereof.

2. A substantially purified variant having at least 90% amino acid sequence identity to the amino acid sequence of claim 1.

3. An isolated and purified polynucleotide encoding the polypeptide of claim 1.

4. An isolated and purified polynucleotide variant having at least 70% polynucleotide sequence identity to the polynucleotide of claim 3.

5. An isolated and purified polynucleotide which hybridizes under stringent conditions to the polynucleotide of claim 3.

6. An isolated and purified polynucleotide having a sequence which is complementary to the polynucleotide of claim 3.

7. A method for detecting a polynucleotide, the method comprising the steps of:

(a) hybridizing the polynucleotide of claim 6 to at least one nucleic acid in a sample, thereby forming a hybridization complex; and

(b) detecting the hybridization complex, wherein the presence of the hybridization complex correlates with the presence of the polynucleotide in the sample.

8. The method of claim 7 further comprising amplifying the polynucleotide prior to hybridization.

9. An isolated and purified polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10, and SEQ ID NO:12, and fragments thereof.

10. An isolated and purified polynucleotide variant having at least 70% polynucleotide sequence identity to the polynucleotide of claim 9.

11. An isolated and purified polynucleotide having a sequence which is complementary to the polynucleotide of claim 9.

12. An expression vector comprising at least a fragment of the polynucleotide of claim 3.

13. A host cell comprising the expression vector of claim 12.

14. A method for producing a polypeptide, the method comprising the steps of:

a) culturing the host cell of claim 13 under conditions suitable for the expression of the polypeptide; and

b) recovering the polypeptide from the host cell culture.

15. A pharmaceutical composition comprising the polypeptide of claim 1 in conjunction with a suitable pharmaceutical carrier.

16. A purified antibody which specifically binds to the polypeptide of claim 1.

17. A purified agonist of the polypeptide of claim 1.

18. A purified antagonist of the polypeptide of claim 1.

19. A method for treating or preventing a disorder associated with decreased expression or activity of HGPRP, the method comprising administering to a subject in need of such treatment an effective amount of the pharmaceutical composition of claim 15.

20. A method for treating or preventing a disorder associated with increased expression or activity of HGPRP, the method comprising administering to a subject in need of such treatment an effective amount of the antagonist of claim 18.

SEQUENCE LISTING

<110> INCYTE PHARMACEUTICALS, INC.
BANDMAN, Olga
LAL, Preeti
TANG, Y. Tom
CORLEY, Neil C.
GUEGLER, Karl J.
GORGONE, Gina A.
BAUGHN, Mariah R.

<120> HUMAN GPCR PROTEINS

<130> PF-0597 PCT

<140> To Be Assigned
<141> Herewith

<150> 09/156,513
<151> 1998-09-17

<160> 12

<170> PERL Program

<210> 1
<211> 441
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1258981CD1

<400> 1
Met Ala Ile His Lys Ala Leu Val Met Cys Leu Gly Leu Pro Leu
1 5 10 15
Phe Leu Phe Pro Gly Ala Trp Ala Gln Gly His Val Pro Pro Gly
20 25 30
Cys Ser Gln Gly Leu Asn Pro Leu Tyr Tyr Asn Leu Cys Asp Arg
35 40 45
Ser Gly Ala Trp Gly Ile Val Leu Glu Ala Val Ala Gly Ala Gly
50 55 60
Ile Val Thr Thr Phe Val Leu Thr Ile Ile Leu Val Ala Ser Leu
65 70 75
Pro Phe Val Gln Asp Thr Lys Lys Arg Ser Leu Leu Gly Thr Gln
80 85 90
Val Phe Phe Leu Leu Gly Thr Leu Gly Leu Phe Cys Leu Val Phe
95 100 105
Ala Cys Val Val Lys Pro Asp Phe Ser Thr Cys Ala Ser Arg Arg
110 115 120
Phe Leu Phe Gly Val Leu Phe Ala Ile Cys Phe Ser Cys Leu Ala
125 130 135
Ala His Val Phe Ala Leu Asn Phe Leu Ala Arg Lys Asn His Gly
140 145 150
Pro Arg Gly Trp Val Ile Phe Thr Val Ala Leu Leu Leu Thr Leu

```
                    155                 160                 165
Val Glu Val Ile Ile Asn Thr Glu Trp Leu Ile Ile Thr Leu Val
                    170                 175                 180
Arg Gly Ser Gly Glu Gly Gly Pro Gln Gly Asn Ser Ser Ala Gly
                    185                 190                 195
Trp Ala Val Ala Ser Pro Cys Ala Ile Ala Asn Met Asp Phe Val
                    200                 205                 210
Met Ala Leu Ile Tyr Val Met Leu Leu Leu Leu Gly Ala Phe Leu
                    215                 220                 225
Gly Ala Trp Pro Ala Leu Cys Gly Arg Tyr Lys Arg Trp Arg Lys
                    230                 235                 240
His Gly Val Phe Val Leu Leu Thr Thr Ala Thr Ser Val Ala Ile
                    245                 250                 255
Trp Val Val Trp Ile Val Met Tyr Thr Tyr Gly Asn Lys Gln His
                    260                 265                 270
Asn Ser Pro Thr Trp Asp Asp Pro Thr Leu Ala Ile Ala Leu Ala
                    275                 280                 285
Ala Asn Ala Trp Ala Phe Val Leu Phe Tyr Val Ile Pro Glu Val
                    290                 295                 300
Ser Gln Val Thr Lys Ser Ser Pro Glu Gln Ser Tyr Gln Gly Asp
                    305                 310                 315
Met Tyr Pro Thr Arg Gly Val Gly Tyr Glu Thr Ile Leu Lys Glu
                    320                 325                 330
Gln Lys Gly Gln Ser Met Phe Val Glu Asn Lys Ala Phe Ser Met
                    335                 340                 345
Asp Glu Pro Val Ala Ala Lys Arg Pro Val Ser Pro Tyr Ser Gly
                    350                 355                 360
Tyr Asn Gly Gln Leu Leu Thr Ser Val Tyr Gln Pro Thr Glu Met
                    365                 370                 375
Ala Leu Met His Lys Val Pro Ser Glu Gly Ala Tyr Asp Ile Ile
                    380                 385                 390
Leu Pro Arg Ala Thr Ala Asn Ser Gln Val Met Gly Ser Ala Asn
                    395                 400                 405
Ser Thr Leu Arg Ala Glu Asp Met Tyr Ser Ala Gln Ser His Gln
                    410                 415                 420
Ala Ala Thr Pro Pro Lys Asp Gly Lys Asn Ser Gln Val Phe Arg
                    425                 430                 435
Asn Pro Tyr Val Trp Asp
                    440

<210> 2
<211> 353
<212> PRT
<213> Homo sapiens


<220>
<221> misc_feature
<223> Incyte ID No: 1459432CD1

<400> 2
Met Asp Leu Glu Ala Ser Leu Leu Pro Thr Gly Pro Asn Ala Ser
 1               5                  10                  15
Asn Thr Ser Asp Gly Pro Asp Asn Leu Thr Ser Ala Gly Ser Pro
                20                  25                  30
Pro Arg Thr Gly Ser Ile Ser Tyr Ile Asn Ile Ile Met Pro Ser
                35                  40                  45
```

Val Phe Gly Thr Ile Cys Leu Leu Gly Ile Ile Gly Asn Ser Thr
50 55 60
Val Ile Phe Ala Val Val Lys Lys Ser Lys Leu His Trp Cys Asn
65 70 75
Asn Val Pro Asp Ile Phe Ile Ile Asn Leu Ser Val Val Asp Leu
80 85 90
Leu Phe Leu Leu Gly Met Pro Phe Met Ile His Gln Leu Met Gly
95 100 105
Asn Gly Val Trp His Phe Gly Glu Thr Met Cys Thr Leu Ile Thr
110 115 120
Ala Met Asp Ala Asn Ser Gln Phe Thr Ser Thr Tyr Ile Leu Thr
125 130 135
Ala Met Ala Ile Asp Arg Tyr Leu Ala Thr Val His Pro Ile Ser
140 145 150
Ser Thr Lys Phe Arg Lys Pro Ser Val Ala Thr Leu Val Ile Cys
155 160 165
Leu Leu Trp Ala Leu Ser Phe Ile Ser Ile Thr Pro Val Trp Leu
170 175 180
Tyr Ala Arg Leu Ile Pro Phe Pro Gly Gly Ala Val Gly Cys Gly
185 190 195
Ile Arg Leu Pro Asn Pro Asp Thr Asp Leu Tyr Trp Phe Thr Leu
200 205 210
Tyr Gln Phe Phe Leu Ala Phe Ala Leu Pro Phe Val Val Ile Thr
215 220 225
Ala Ala Tyr Val Arg Ile Leu Gln Arg Met Thr Ser Ser Val Ala
230 235 240
Pro Thr Ser Gln Arg Ser Ile Arg Leu Arg Thr Lys Arg Val Thr
245 250 255
Arg Thr Ala Ile Ala Ile Cys Leu Val Phe Phe Val Cys Trp Ala
260 265 270
Pro Tyr Tyr Val Leu Gln Leu Thr Gln Leu Ser Ile Ser Arg Pro
275 280 285
Thr Pro Thr Phe Val Tyr Leu Tyr Asn Ala Ala Ile Ser Leu Gly
290 295 300
Tyr Ala Asn Ser Cys Leu Asn Pro Phe Val Tyr Ile Val Leu Cys
305 310 315
Glu Thr Phe Arg Lys Arg Leu Val Leu Ser Val Lys Pro Ala Ala
320 325 330
Gln Gly Gln Leu Arg Ala Val Ser Asn Ala Gln Ala Ala Asp Glu
335 340 345
Glu Arg Thr Glu Ser Lys Gly Thr
350

<210> 3
<211> 333
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2214673CD1

<400> 3
Met Trp Ser Cys Ser Trp Phe Asn Gly Thr Gly Leu Val Glu Glu
1 5 10 15
Leu Pro Ala Cys Gln Asp Leu Gln Leu Gly Leu Ser Leu Leu Ser
20 25 30

Leu Leu Gly Leu Val Val Gly Val Pro Val Gly Leu Cys Tyr Asn
35 40 45
Ala Leu Leu Val Leu Ala Asn Leu His Ser Lys Ala Ser Met Thr
50 55 60
Met Pro Asp Val Tyr Phe Val Asn Met Ala Val Ala Gly Leu Val
65 70 75
Leu Ser Ala Leu Ala Pro Val His Leu Leu Gly Pro Pro Ser Ser
80 85 90
Arg Trp Ala Leu Trp Ser Val Gly Gly Glu Val His Val Ala Leu
95 100 105
Gln Ile Pro Phe Asn Val Ser Ser Leu Val Ala Met Tyr Ser Thr
110 115 120
Ala Leu Leu Ser Leu Asp His Tyr Ile Glu Arg Ala Leu Pro Arg
125 130 135
Thr Tyr Met Ala Ser Val Tyr Asn Thr Arg His Val Cys Gly Phe
140 145 150
Val Trp Gly Gly Ala Leu Leu Thr Ser Phe Ser Ser Leu Leu Phe
155 160 165
Tyr Ile Cys Ser His Val Ser Thr Arg Ala Leu Glu Cys Ala Lys
170 175 180
Met Gln Asn Ala Glu Ala Ala Asp Ala Thr Leu Val Phe Ile Gly
185 190 195
Tyr Val Val Pro Ala Leu Ala Thr Leu Tyr Ala Leu Val Leu Leu
200 205 210
Ser Arg Val Arg Arg Glu Asp Thr Pro Leu Asp Arg Asp Thr Gly
215 220 225
Arg Leu Glu Pro Ser Ala His Arg Leu Leu Val Ala Thr Val Cys
230 235 240
Thr Gln Phe Gly Leu Trp Thr Pro His Tyr Leu Ile Leu Leu Gly
245 250 255
His Thr Gly Ile Ile Ser Arg Gly Lys Pro Val Asp Ala His Tyr
260 265 270
Leu Gly Leu Leu His Phe Val Lys Asp Phe Ser Lys Leu Leu Ala
275 280 285
Phe Ser Ser Ser Phe Val Thr Pro Leu Leu Tyr Arg Tyr Met Asn
290 295 300
Gln Ser Phe Pro Ser Lys Leu Gln Arg Leu Met Lys Lys Leu Pro
305 310 315
Cys Gly Asp Arg His Cys Ser Pro Asp His Met Gly Val Gln Gln
320 325 330
Val Leu Ala

<210> 4
<211> 396
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2488822CD1

<400> 4
Met Phe Arg Pro Leu Val Asn Leu Ser His Ile Tyr Phe Lys Lys
1 5 10 15
Phe Gln Tyr Cys Gly Tyr Ala Pro His Val Arg Ser Cys Lys Pro
20 25 30

Asn Thr Asp Gly Ile Ser Ser Leu Glu Asn Leu Leu Ala Ser Ile
35 40 45
Ile Gln Arg Val Phe Val Trp Val Val Ser Ala Val Thr Cys Phe
50 55 60
Gly Asn Ile Phe Val Ile Cys Met Arg Pro Tyr Ile Arg Ser Glu
65 70 75
Asn Lys Leu Tyr Ala Met Ser Ile Ile Ser Leu Cys Cys Ala Asp
80 85 90
Cys Leu Met Gly Ile Tyr Leu Phe Val Ile Gly Gly Phe Asp Leu
95 100 105
Lys Phe Arg Gly Glu Tyr Asn Lys His Ala Gln Leu Trp Met Glu
110 115 120
Ser Thr His Cys Gln Leu Val Gly Ser Leu Ala Ile Leu Ser Thr
125 130 135
Glu Val Ser Val Leu Leu Leu Thr Phe Leu Thr Leu Glu Lys Tyr
140 145 150
Ile Cys Ile Val Tyr Pro Phe Arg Cys Val Arg Pro Gly Lys Cys
155 160 165
Arg Thr Ile Thr Val Leu Ile Leu Ile Trp Ile Thr Gly Phe Ile
170 175 180
Val Ala Phe Ile Pro Leu Ser Asn Lys Glu Phe Phe Lys Asn Tyr
185 190 195
Tyr Ala Pro Asn Gly Val Cys Phe Pro Leu His Ser Glu Asp Thr
200 205 210
Glu Ser Ile Gly Ala Gln Ile Tyr Ser Val Ala Ile Phe Leu Gly
215 220 225
Ile Asn Leu Ala Ala Phe Ile Ile Ile Val Phe Ser Tyr Gly Ser
230 235 240
Met Phe Tyr Ser Val His Gln Ser Ala Ile Thr Ala Thr Glu Ile
245 250 255
Arg Asn Gln Val Lys Lys Glu Met Ile Leu Ala Lys Arg Phe Phe
260 265 270
Phe Ile Val Phe Thr Asp Ala Leu Cys Trp Ile Pro Ile Phe Val
275 280 285
Val Lys Phe Leu Ser Leu Leu Gln Val Glu Ile Pro Gly Thr Ile
290 295 300
Thr Ser Trp Val Val Ile Phe Ile Leu Pro Ile Asn Ser Ala Leu
305 310 315
Asn Pro Ile Leu Tyr Thr Leu Thr Thr Arg Pro Phe Lys Glu Met
320 325 330
Ile His Arg Phe Trp Tyr Asn Tyr Arg Gln Arg Lys Ser Met Asp
335 340 345
Ser Lys Gly Gln Lys Thr Tyr Ala Pro Ser Phe Ile Trp Val Glu
350 355 360
Met Trp Pro Leu Gln Glu Met Pro Pro Glu Leu Met Lys Pro Asp
365 370 375
Leu Phe Thr Tyr Pro Cys Glu Met Ser Leu Ile Ser Gln Ser Thr
380 385 390
Arg Leu Asn Ser Tyr Ser
395

<210> 5
<211> 403
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2705201CD1

<400> 5
Met Phe Val Ala Ser Glu Arg Lys Met Arg Ala His Gln Val Leu
1 5 10 15
Thr Phe Leu Leu Leu Phe Val Ile Thr Ser Val Ala Ser Glu Asn
20 25 30
Ala Ser Thr Ser Arg Gly Cys Gly Leu Asp Leu Leu Pro Gln Tyr
35 40 45
Val Ser Leu Cys Asp Leu Asp Ala Ile Trp Gly Ile Val Val Glu
50 55 60
Ala Val Ala Gly Ala Gly Ala Leu Ile Thr Leu Leu Leu Met Leu
65 70 75
Ile Leu Leu Val Arg Leu Pro Phe Ile Lys Glu Lys Glu Lys Lys
80 85 90
Ser Pro Val Gly Leu His Phe Leu Phe Leu Leu Gly Thr Leu Gly
95 100 105
Leu Phe Gly Leu Thr Phe Ala Phe Ile Ile Gln Glu Asp Glu Thr
110 115 120
Ile Cys Ser Val Arg Arg Phe Leu Trp Gly Val Leu Phe Ala Leu
125 130 135
Cys Phe Ser Cys Leu Leu Ser Gln Ala Trp Arg Val Arg Arg Leu
140 145 150
Val Arg His Gly Thr Gly Pro Ala Gly Trp Gln Leu Val Gly Leu
155 160 165
Ala Leu Cys Leu Met Leu Val Gln Val Ile Ile Ala Val Glu Trp
170 175 180
Leu Val Leu Thr Val Leu Arg Asp Thr Arg Pro Ala Cys Ala Tyr
185 190 195
Glu Pro Met Asp Phe Val Met Ala Leu Ile Tyr Asp Met Val Leu
200 205 210
Leu Val Val Thr Leu Gly Leu Ala Leu Phe Thr Leu Cys Gly Lys
215 220 225
Phe Lys Arg Trp Lys Leu Asn Gly Ala Phe Leu Leu Ile Thr Ala
230 235 240
Phe Leu Ser Val Leu Ile Trp Val Ala Trp Met Thr Met Tyr Leu
245 250 255
Phe Gly Asn Val Lys Leu Gln Gln Gly Asp Ala Trp Asn Asp Pro
260 265 270
Thr Leu Ala Ile Thr Leu Ala Ala Ser Gly Trp Val Phe Val Ile
275 280 285
Phe His Ala Ile Pro Glu Ile His Cys Thr Leu Leu Pro Ala Leu
290 295 300
Gln Glu Asn Thr Pro Asn Tyr Phe Asp Thr Ser Gln Pro Arg Met
305 310 315
Arg Glu Thr Ala Phe Glu Glu Asp Val Gln Leu Pro Arg Ala Tyr
320 325 330
Met Glu Asn Lys Ala Phe Ser Met Asp Glu His Asn Ala Ala Leu
335 340 345
Arg Thr Ala Gly Phe Pro Asn Gly Ser Leu Gly Lys Arg Pro Ser
350 355 360
Gly Ser Leu Gly Lys Arg Pro Ser Ala Pro Phe Arg Ser Asn Val
365 370 375
Tyr Gln Pro Thr Glu Met Ala Val Val Leu Asn Gly Gly Thr Ile
380 385 390

```
Pro Thr Ala Pro Pro Ser His Thr Gly Arg His Leu Trp
                395                 400

<210> 6
<211> 807
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 3036563CD1

<400> 6
Met Gly Thr Tyr His Cys Ile Phe Arg Tyr Lys Asn Ser Tyr Ser
  1               5                  10                  15
Ile Ala Thr Lys Asp Val Ile Val His Pro Leu Pro Leu Lys Leu
                 20                  25                  30
Asn Ile Met Val Asp Pro Leu Glu Ala Thr Val Ser Cys Ser Gly
                 35                  40                  45
Ser His His Ile Lys Cys Cys Ile Glu Glu Asp Gly Asp Tyr Lys
                 50                  55                  60
Val Thr Phe His Met Gly Ser Ser Ser Leu Pro Ala Ala Lys Glu
                 65                  70                  75
Val Asn Lys Lys Gln Val Cys Tyr Lys His Asn Phe Asn Ala Ser
                 80                  85                  90
Ser Val Ser Trp Cys Ser Lys Thr Val Asp Val Cys Cys His Phe
                 95                 100                 105
Thr Asn Ala Ala Asn Asn Ser Val Trp Ser Pro Ser Met Lys Leu
                110                 115                 120
Asn Leu Val Pro Gly Glu Asn Ile Thr Cys Gln Asp Pro Val Ile
                125                 130                 135
Gly Val Gly Glu Pro Gly Lys Val Ile Gln Lys Leu Cys Arg Phe
                140                 145                 150
Ser Asn Val Pro Ser Ser Pro Glu Ser Pro Ile Gly Gly Thr Ile
                155                 160                 165
Thr Tyr Lys Cys Val Gly Ser Gln Trp Glu Glu Lys Arg Asn Asp
                170                 175                 180
Cys Ile Ser Ala Pro Ile Asn Ser Leu Leu Gln Met Ala Lys Ala
                185                 190                 195
Leu Ile Lys Ser Pro Ser Gln Asp Glu Met Leu Pro Thr Tyr Leu
                200                 205                 210
Lys Asp Leu Ser Ile Ser Ile Gly Lys Ala Glu His Glu Ile Ser
                215                 220                 225
Ser Ser Pro Gly Ser Leu Gly Ala Ile Ile Asn Ile Leu Asp Leu
                230                 235                 240
Leu Ser Thr Val Pro Thr Gln Val Asn Ser Glu Met Met Thr His
                245                 250                 255
Val Leu Ser Thr Val Asn Ile Ile Leu Gly Lys Pro Val Leu Asn
                260                 265                 270
Thr Trp Lys Val Leu Gln Gln Gln Trp Thr Asn Gln Ser Ser Gln
                275                 280                 285
Leu Leu His Ser Val Glu Arg Phe Ser Gln Ala Leu Gln Ser Gly
                290                 295                 300
Asp Ser Pro Pro Leu Ser Phe Ser Gln Thr Asn Val Gln Met Ser
                305                 310                 315
Ser Met Val Ile Lys Ser Ser His Pro Glu Thr Tyr Gln Gln Arg
                320                 325                 330
```

Phe Val Phe Pro Tyr Phe Asp Leu Trp Gly Asn Val Val Ile Asp
335 340 345
Lys Ser Tyr Leu Glu Asn Leu Gln Ser Asp Ser Ser Ile Val Thr
350 355 360
Met Ala Phe Pro Thr Leu Gln Ala Ile Leu Ala Gln Asp Ile Gln
365 370 375
Glu Asn Asn Phe Ala Glu Ser Leu Val Met Thr Thr Thr Val Ser
380 385 390
His Asn Thr Thr Met Pro Phe Arg Ile Ser Met Thr Phe Lys Asn
395 400 405
Asn Ser Pro Ser Gly Gly Glu Thr Lys Cys Val Phe Trp Asn Phe
410 415 420
Arg Leu Ala Asn Asn Thr Gly Gly Trp Asp Ser Ser Gly Cys Tyr
425 430 435
Val Glu Glu Gly Asp Gly Asp Asn Val Thr Cys Ile Cys Asp His
440 445 450
Leu Thr Ser Phe Ser Ile Leu Met Ser Pro Asp Ser Pro Asp Pro
455 460 465
Ser Ser Leu Leu Gly Ile Leu Leu Asp Ile Ile Ser Tyr Val Gly
470 475 480
Val Gly Phe Ser Ile Leu Ser Leu Ala Ala Cys Leu Val Val Glu
485 490 495
Ala Val Val Trp Lys Ser Val Thr Lys Asn Arg Thr Ser Tyr Met
500 505 510
Arg His Thr Cys Ile Val Asn Ile Ala Ala Ser Leu Leu Val Ala
515 520 525
Asn Thr Trp Phe Ile Val Val Ala Ala Ile Gln Asp Asn Arg Tyr
530 535 540
Ile Leu Cys Lys Thr Ala Cys Val Ala Ala Thr Phe Phe Ile His
545 550 555
Phe Phe Tyr Leu Ser Val Phe Phe Trp Met Leu Thr Leu Gly Leu
560 565 570
Met Leu Phe Tyr Arg Leu Val Phe Ile Leu His Glu Thr Ser Arg
575 580 585
Ser Thr Gln Lys Ala Ile Ala Phe Cys Leu Gly Tyr Gly Cys Pro
590 595 600
Leu Ala Ile Ser Val Ile Thr Leu Gly Ala Thr Gln Pro Arg Glu
605 610 615
Val Tyr Thr Arg Lys Asn Val Cys Trp Leu Asn Trp Glu Asp Thr
620 625 630
Lys Ala Leu Leu Ala Phe Ala Ile Pro Ala Leu Ile Ile Val Val
635 640 645
Val Asn Ile Thr Ile Thr Ile Val Val Ile Thr Lys Ile Leu Arg
650 655 660
Pro Ser Ile Gly Asp Lys Pro Cys Lys Gln Glu Lys Ser Ser Leu
665 670 675
Phe Gln Ile Ser Lys Ser Ile Gly Val Leu Thr Pro Leu Leu Gly
680 685 690
Leu Thr Trp Gly Phe Gly Leu Thr Thr Val Phe Pro Gly Thr Asn
695 700 705
Leu Val Phe His Ile Ile Phe Ala Ile Leu Asn Val Phe Gln Gly
710 715 720
Leu Phe Ile Leu Leu Phe Gly Cys Leu Trp Asp Leu Lys Val Gln
725 730 735
Glu Ala Leu Leu Asn Lys Phe Ser Leu Ser Arg Trp Ser Ser Gln
740 745 750
His Ser Lys Ser Thr Ser Leu Gly Ser Ser Thr Pro Val Phe Ser

```
                755                 760                 765
Met Ser Ser Pro Ile Ser Arg Arg Phe Asn Asn Leu Phe Gly Lys
                770                 775                 780
Thr Gly Thr Tyr Asn Val Ser Thr Pro Glu Ala Thr Ser Ser Ser
                785                 790                 795
Leu Glu Asn Ser Ser Ser Ala Ser Ser Leu Leu Asn
                800                 805
```

<210> 7
<211> 1824
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1258981CB1

<400> 7

```
cggctcgagc cctcaccagc cggaaagtac gagtcggctc agcctggagg gacccaacca   60
gagcctggcc tgggagccag gatggccatc cacaaagcct tggtgatgtg cctgggactg  120
cctctcttcc tgttcccagg ggcctgggcc cagggccatg tcccacccgg ctgcagccaa  180
ggcctcaacc ccctgtacta caacctgtgt gaccgctctg gggcgtgggg catcgtcctg  240
gaggccgtgg ctggggcggg cattgtcacc acgtttgtgc tcaccatcat cctggtggcc  300
agcctcccct ttgtgcagga caccaagaaa cggagcctgc tggggaccca ggtattcttc  360
cttctgggga ccctgggcct cttctgcctc gtgtttgcct gtgtggtgaa gcccgacttc  420
tccacctgtg cctctcggcg cttcctcttt ggggttctgt tcgccatctg cttctcttgt  480
ctggcggctc acgtctttgc cctcaacttc ctggcccgga agaaccacgg gccccggggc  540
tgggtgatct tcactgtggc tctgctgctg accctggtag aggtcatcat caatacagag  600
tggctgatca tcaccctggt tcggggcagt ggcgagggcg gccctcaggg caacagcagc  660
gcaggctggg ccgtggcctc cccctgtgcc atcgccaaca tggactttgt catggcactc  720
atctacgtca tgctgctgct gctgggtgcc ttcctggggg cctggcccgc cctgtgtggc  780
cgctacaagc gctggcgtaa gcatggggtc tttgtgctcc tcaccacagc cacctccgtt  840
gccatatggg tggtgtggat cgtcatgtat acttacggca acaagcagca caacagtccc  900
acctgggatg accccacgct ggccatcgcc ctcgccgcca atgcctgggc cttcgtcctc  960
ttctacgtca tccccgaggt ctcccaggtg accaagtcca gcccagagca aagctaccag 1020
ggggacatgt accccacccg gggcgtgggc tatgagacca tcctgaaaga gcagaagggt 1080
cagagcatgt tcgtggagaa caaggccttt tccatggatg agccggttgc agctaagagg 1140
ccggtgtcac catacagcgg gtacaatggg cagctgctga ccagtgtgta ccagcccact 1200
gagatggccc tgatgcacaa agttccgtcc gaaggagctt acgacatcat cctcccacgg 1260
gccaccgcca acagccaggt gatgggcagt gccaactcga ccctgcgggc tgaagacatg 1320
tactcggccc agagccacca ggcggccaca ccgccgaaag acggcaagaa ctctcaggtc 1380
tttagaaacc cctacgtgtg ggactgagtc agcggtggcg aggagaggcg gtcggatttg 1440
gggagggccc tgaggacctg gccccgggca agggactctc caggctcctc ctccccctgg 1500
caggcccagc aacatgtgcc ccagatgtgg aagggcctcc ctctctgcca gtgtttgggt 1560
gggtgtcatg ggtgtcccca cccactcctc agtgtttgtg gagtcgagga gccaacccca 1620
gcctcctgcc aggatcacct cggcggtcac actccagcca aatagtgttc tcggggtggt 1680
ggctgggcag cgcctatgtt tctctggaga ttcctgcaac ctcaagagac ttcccaggcg 1740
ctcaggcctg gatcttgctc ctctgtgagg aacaagggtg cctaataaat acatttctgc 1800
tttattaact cttaaaaaaa aaaa                                       1824
```

<210> 8
<211> 2152
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1459432CB1

<400> 8
ttatgtctgg tcgactctga attgggcttg gaggcggcac ggctgccagg ctacggaggt 60
agaccccctt cccaactgcg gggcttgcgc tccgggacaa ggtggcaggc gctggaggct 120
gccgcagcct gcgtgggtgg aggggagctc agctcggttg tggcagcatg cgaccggcac 180
tggctggatg gacctggaag cctcgctgct gcccactggt cccaatgcca gcaacacctc 240
tgatggcccc gataacctca cttcggcagg atcacctcct cgcacgggga gcatctccta 300
catcaacatc atcatgcctt cggtgttcgg caccatctgc ctcctgggca tcatcgggaa 360
ctccacggtc atcttcgcgg tcgtgaagaa gtccaagctg cactggtgca acaacgtccc 420
cgacatcttc atcatcaacc tctcggtagt agatctcctc tttctcctgg gcatgccctt 480
catgatccac cagctcatgg gcaatggggt gtggcacttt ggggagacca tgtgcaccct 540
catcacggcc atggatgcca atagtcagtt caccagcacc tacatcctga ccgccatggc 600
cattgaccgc tacctggcca ctgtccaccc catctcttcc acgaagttcc ggaagccctc 660
tgtggccacc ctggtgatct gcctcctgtg ggccctctcc ttcatcagca tcacccctgt 720
gtggctgtat gccagactca tccccttccc aggaggtgca gtgggctgcg gcatacgcct 780
gcccaaccca gacactgacc tctactggtt caccctgtac cagtttttcc tggcctttgc 840
cctgcctttt gtggtcatca cagccgcata cgtgaggatc ctgcagcgca tgacgtcctc 900
agtggccccc acctcccagc gcagcatccg gctgcggaca aagagggtga cccgcacagc 960
catcgccatc tgtctggtct tctttgtgtg ctgggcaccc tactatgtgc tacagctgac 1020
ccagttgtcc atcagccgcc cgacccccac ctttgtctac ttatacaatg cggccatcag 1080
cttgggctat gccaacagct gcctcaaccc gtttgtgtac atcgtgctct gtgagacgtt 1140
ccgcaaacgc ttggtcctgt cggtgaagcc tgcagcccag gggcagcttc gcgctgtcag 1200
caacgctcag gcggctgacg aggagaggac agaaagcaaa ggcacctgat acttcccctg 1260
ccaccctgca cacctccaag tcagggcacc acaacacgcc accgggagag atgctgagaa 1320
aaacccaaga ccgctcggga aatgcaggaa ggccgggttg tgaggggttg ttgcaatgaa 1380
ataaatacat tccatggggc tcacacgttg ctggggaggc ctggagtcag gtttggggtt 1440
ttcagatatc agaaatcccc ttgggggagc aggatgagac ctttggatag aacagaagct 1500
gagcaagaga acatgttggt ttggataacc ggttgcacta tatctgtgag ctctcaaatg 1560
tcttcttccc aaggcaagag gtggaagggt actgactggg tttgtttaaa gtcaggcagg 1620
gctggagtga gcagccaggg ccatgttgca caaggcctga gagacgggaa agggcccgat 1680
cgctctttcc cgcctctcac tggtgcgatg gaaggtggcc tttctcccaa gctggtggat 1740
aatgaaaaat aaagcatccc atctctcggc gttccagcat cctgtcaatt tccctttttgc 1800
tctagaggat gcatgtttat ttgaggggat gtggcactga gcccacagga gtaaaagccc 1860
agtttgctag gaggtctgct tactgaaaac aaggagacct ggggtgggtg tggttggggg 1920
tcttaaaact aataaaagct ggggtcgggg ggcttttgca gctctggtga cattctctcc 1980
acggggcaca tttgctcagt cactaatcca gcttgagtgt ccgtgtgttc tgcatgtgca 2040
ggggtcattc tagtgcccgg tgtgttggca tcatcttttt gctctagccc ttcctctcca 2100
aaataaaatc aaataaagga aaatctccac ccaaaaaaaa aaaaaaaaaa gg 2152

<210> 9
<211> 1878
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2214673CB1

<400> 9
cgcacagcgc gcaggtcctc accagagctc tggtggccac ctctgtcccg ccatgctgct 60
caccgacagt ggccagggcc cacagcacca agaggcttgg gccacaaagt aaagggtcgc 120
ggagcctcgc cggccgccat gtggagctgc agctggttca acggcacagg gctggtggag 180
gagctgcctg cctgccagga cctgcagctg gggctgtcac tgttgtcgct gctgggcctg 240
gtggtgggcg tgccagtggg cctgtgctac aacgccctgc tggtgctggc caacctacac 300

agcaaggcca gcatgaccat gccggacgtg tactttgtca acatggcagt ggcaggcctg 360
gtgctcagcg ccctggcccc tgtgcacctg ctcggccccc cgagctcccg gtgggcgctg 420
tggagtgtgg gcggcgaagt ccacgtggca ctgcagatcc ccttcaatgt gtcctcactg 480
gtggccatgt actccaccgc cctgctgagc ctcgaccact acatcgagcg tgcactgccg 540
cggacctaca tggccagcgt gtacaacacg cggcacgtgt gcggcttcgt gtggggtggc 600
gcgctgctga ccagcttctc ctcgctgctc ttctacatct gcagccatgt gtccacccgc 660
gcgctagagt gcgccaagat gcagaacgca gaagctgccg acgccacgct ggtgttcatc 720
ggctacgtgg tgccagcact ggccaccctc tacgcgctgg tgctactctc ccgcgtccgc 780
agggaggaca cgcccctgga ccgggacacg ggccggctgg agccctcggc acacaggctg 840
ctggtggcca ccgtgtgcac gcagtttggg ctctggacgc cacactatct gatcctgctg 900
gggcacacgg gcatcatctc gcgagggaag cccgtggacg cacactacct ggggctactg 960
cactttgtga aggatttctc caaactcctg gccttctcca gcagctttgt gacaccactt 1020
ctctaccgct acatgaacca gagcttcccc agcaagctcc aacggctgat gaaaaagctg 1080
ccctgcgggg accggcactg ctccccggac cacatggggg tgcagcaggt gctggcgtag 1140
gcggcccagc cctcctgggg agacgtgact ctggtggacg cagagcactt agttaccctg 1200
gacgctcccc acatccttcc agaaggagac gagctgctgg aagagaagca ggaggggtgt 1260
ttttcttgaa gtttcctttt tcccacaaat gccactcttg ggccaaggct gtggtccccg 1320
tggctggcat ctggcttgag tctccccgag gcctgtgcgt ctcccaaaca cgcagctcaa 1380
ggtccacatc cgcaaaagcc tcctcgcctt cagcctcctc agcattcagt ttgtcaatga 1440
agtgatgaaa gcttagagcc agtatttata ctttgtggtt aaaatacttg attcccccctt 1500
gtttgtttta caaaaacaga tgtttcctag aaaaatgaca aatagtaaaa tgaacaaaac 1560
cctacgaaag aatggcaaca gccagggtgg ccgggccctg ccagtgggcg gcgtgtgcta 1620
gcaaggcctg ccgggtgtgc cgcagtcacc acagggttct gagaacattt cacagaagtg 1680
cctgagacgc ggagacatgg ctggtgttaa atggagctat tcaatagcag tgacgcgctc 1740
tcctcagcca ccaaatgtcc ctgacaccct ccccagcccc cacagataac atcagctgag 1800
gtttttttca gtatgaacct gtcctaaatc aattcctcaa agtgtgcaca aaactaaaga 1860
atataaataa acagaagc 1878

<210> 10
<211> 1804
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2488822CB1

<400> 10
taagtgttaa ctaaaagcat tttattaaat tgtccttcac agaaactcaa tttattaaac 60
catgtataat acatgttcct ttgattgatt attaatttga tatttttagc agcctagaag 120
ggattgaaat ttcaaatatc caacaaagga tgtttagacc tcttgtgaat ctctctcaca 180
tatattttaa gaaattccag tactgtgggt atgcaccaca tgttcgcagc tgtaaaccaa 240
acactgatgg aatttcatct ctagagaatc tcttggcaag cattattcag agagtatttg 300
tctgggttgt atctgcagtt acctgctttg gaaacatttt tgtcatttgc atgcgacctt 360
atatcaggtc tgagaacaag ctgtatgcca tgtcaatcat ttctctctgc tgtgccgact 420
gcttaatggg aatatattta ttcgtgatcg gaggctttga cctaaagttt cgtggagaat 480
acaataagca tgcgcagctg tggatggaga gtactcattg tcagcttgta ggatctttgg 540
ccattctgtc cacagaagta tcagttttac tgttaacatt tctgacattg gaaaaataca 600
tctgcattgt ctatcctttt agatgtgtga gacctggaaa atgcagaaca attacagttc 660
tgattctcat ttggattact ggttttatag tggctttcat tccattgagc aataaggaat 720
ttttcaaaaa ctactatgca cccaatggag tatgcttccc tcttcattca gaagatacag 780
aaagtattgg agcccagatt tattcagtgg caatttttct tggtattaat ttggccgcat 840
ttatcatcat agtttttttcc tatggaagca tgttttatag tgttcatcaa agtgccataa 900
cagcaactga aatacggaat caagttaaaa aagagatgat ccttgccaaa cgttttttct 960
ttatagtatt tactgatgca ttatgctgga tacccatttt tgtagtgaaa tttctttcac 1020
tgcttcaggt agaaatacca ggtaccataa cctcttgggt agtgattttt attctgccca 1080
ttaacagtgc tttgaaccca attctctata ctctgaccac aagaccattt aaagaaatga 1140

ttcatcggtt ttggtataac tacagacaaa gaaaatctat ggacagcaaa ggtcagaaaa 1200
catatgctcc atcattcatc tgggtggaaa tgtggccact gcaggagatg ccacctgagt 1260
taatgaagcc ggaccttttc acataccect gtgaaatgtc actgatttct caatcaacga 1320
gactcaattc ctattcatga ctgactctga aattcatttc ttcgcagaga atactgtggg 1380
ggtgcttcat gagggattta ctggtatgaa atgaatacca caaaattaat ttataataat 1440
agctaagata aatattttac aaggacatga ggaaaaataa aaatgactaa tgctcttaca 1500
aagggaagta attatatcaa taatgtatat atattagtag acattttgca taagaaatta 1560
agagaaatct acttcagtaa cattcattca tttttctaac atgcatttat tgagtaccca 1620
ctactatgtg catagcattg caatatagtc ctggaagtag acagtgcaga acctttcaat 1680
ctgtagatgg tgtttaatga caaaagacta tacaaagtcc atctgcagtt cctagtttaa 1740
agtagagctt tacctgtcat gtgcatcagc aagaatcata gcgattttaa atagaggtgt 1800
ggac 1804

<210> 11
<211> 1520
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2705201CB1

<400> 11
tgccgaagag tctggagcgt cggcgctgcg gggccgcggg ggtcgaatgt tcgtggcatc 60
agagagaaag atgagagctc accaggtgct caccttcctc ctgctcttcg tgatcacctc 120
ggtggcctct gaaaacgcca gcacatcccg aggctgtggg ctggacctcc tccctcagta 180
cgtgtccctg tgcgacctgg acgccatctg ggcattgtg gtggaggcgg tggccggggc 240
gggcgccctg atcacactgc tcctgatgct catcctcctg gtgcggctgc ccttcatcaa 300
ggagaaggag aagaagagcc ctgtgggcct ccactttctg ttcctcctgg ggaccctggg 360
cctctttggg ctgacgtttg ccttcatcat ccaggaggac gagaccatct gctctgtccg 420
ccgcttcctc tggggcgtcc tctttgcgct ctgcttctcc tgcctgctga gccaggcatg 480
gcgcgtgcgg aggctggtgc ggcatggcac gggccccgcg ggctggcagc tggtgggcct 540
ggcgctgtgc ctgatgctgg tgcaagtcat catcgctgtg gagtggctgg tgctcaccgt 600
gctgcgtgac acaaggccag cctgcgccta cgagcccatg gactttgtga tggccctcat 660
ctacgacatg gtactgcttg tggtcaccct ggggctggcc ctcttcactc tgtgcggcaa 720
gttcaagagg tggaagctga acggggcctt cctcctcatc acagccttcc tctctgtgct 780
catctgggtg gcctggatga ccatgtacct cttcggcaat gtcaagctgc agcaggggga 840
tgcctggaac gaccccacct tggccatcac gctggcggcc agcggctggg tcttcgtcat 900
cttccacgcc atccctgaga tccactgcac ccttctgcca gccctgcagg agaacacgcc 960
caactacttc gacacgtcgc agcccaggat gcgggagacg gccttcgagg aggacgtgca 1020
gctgccgcgg gcctatatgg agaacaaggc cttctccatg gatgaacaca atgcagctct 1080
ccgaacagca ggatttccca acggcagctt gggaaaaaga cccagtggca gcttggggaa 1140
aagacccagc gctccgttta gaagcaacgt gtatcagcca actgagatgg ccgtcgtgct 1200
caacggtggg accatcccaa ctgctccgcc aagtcacaca ggaagacacc tttggtgaaa 1260
gactttaagt tccagagaat cagaatttct cttaccgatt tgcctccctg gctgtgtctt 1320
tcttgaggga gaaatcggta acagttgccg aaccaggccg cctcacagcc aggaaatttg 1380
gaaatcctag ccaaggggat ttcgtgtaaa tgtgaacact gacgaactga aaagctaaca 1440
ccgactgccc gcccctcccc tgccacacac acagacacgt aataccagac caacctcaat 1500
ccccacctta aaaaaaaaaa 1520

<210> 12
<211> 2919
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 3036563CB1

<400> 12
```
atcttgatgg agcagaatca gtactgacag tcaagacctc gaccagggag tggaatggga   60
acctatcact gcatatttag atataagaat tcatacagta ttgcaaccaa agacgtcatt  120
gttcacccgc tgcctctaaa gctgaacatc atggttgatc ctttggaagc tactgtttca  180
tgcagtggtt cccatcacat caagtgctgc atagaggagg atggagacta caaagttact  240
ttccatatgg gttcctcatc ccttcctgct gcaaaagaag ttaacaaaaa acaagtgtgc  300
tacaaacaca atttcaatgc aagctcagtt tcctggtgtt caaaaactgt tgatgtgtgt  360
tgtcacttta ccaatgctgc taataattca gtctggagcc catctatgaa gctgaatctg  420
gttcctgggg aaaacatcac atgccaggat cccgtaatag gtgtcggaga gccgggggaaa 480
gtcatccaga agctatgccg gttctcaaac gttcccagca gccctgagag tcccattggc  540
gggaccatca cttacaaatg tgtaggctcc cagtgggagg agaagagaaa tgactgcatc  600
tctgccccaa taaacagtct gctccagatg gctaaggctt tgatcaagag cccctctcag  660
gatgagatgc tccctacata cctgaaggat ctttctatta gcataggcaa agcggaacat  720
gaaatcagct cttctcctgg gagtctggga gccattatta acatccttga tctgctctca  780
acagttccaa cccaagtaaa ttcagaaatg atgacgcacg tgctctctac ggttaatatc  840
atccttggca agcccgtctt gaacacctgg aaggttttac aacagcaatg gaccaatcag  900
agttcacagc tactacattc agtggaaaga ttttcccaag cattacagtc aggagatagc  960
cctccattgt ccttctccca aactaatgtg cagatgagca gcatggtaat caagtccagc 1020
cacccagaaa cctatcaaca gaggtttgtt ttcccatact ttgacctctg gggcaatgtg 1080
gtcattgaca agagctacct agaaaacttg cagtcggatt cgtctattgt caccatggct 1140
ttcccaactc tccaagccat ccttgctcag gatatccagg aaaataactt tgcagagagc 1200
ttagtgatga caaccactgt cagccacaat acgactatgc cattcaggat ttcaatgact 1260
tttaagaaca atagcccttc aggcggcgaa acgaagtgtg tcttctggaa cttcaggctt 1320
gccaacaaca caggggggtg ggacagcagt gggtgctatg ttgaagaagg tgatggggac 1380
aatgtcacct gtatctgtga ccacctaaca tcattctcca tcctcatgtc ccctgactcc 1440
ccagatccta gttctctcct gggaatactc ctggatatta tttcttatgt tggggtgggc 1500
ttttccatct tgagcttggc agcctgtcta gttgtggaag ctgtggtgtg gaaatcggtg 1560
accaagaatc ggacttctta tatgcgccac acctgcatag tgaatatcgc tgcctccctt 1620
ctggtcgcca acacctggtt cattgtggtc gctgccatcc aggacaatcg ctacatactc 1680
tgcaagacag cctgtgtggc tgccaccttc ttcatccact tcttctacct cagcgtcttc 1740
ttctggatgc tgacactggg cctcatgctg ttctatcgcc tggttttcat tctgcatgaa 1800
acaagcaggt ccactcagaa agccattgcc ttctgtcttg gctatggctg cccacttgcc 1860
atctcggtca tcacgctggg agccacccag ccccgggaag tctatacgag gaagaatgtc 1920
tgttggctca actgggagga caccaaggcc ctgctggctt tcgccatccc agcactgatc 1980
attgtggtgg tgaacataac catcactatt gtggtcatca ccaagatcct gaggccttcc 2040
attggagaca agccatgcaa gcaggagaag agcagcctgt ttcagatcag caagagcatt 2100
ggggtcctca caccactctt gggcctcact tggggttttg gtctcaccac tgtgttccca 2160
gggaccaacc ttgtgttcca tatcatattt gccatcctca atgtcttcca gggattattc 2220
attttactct ttggatgcct ctgggatctg aaggtacagg aagctttgct gaataagttt 2280
tcattgtcga gatggtcttc acagcactca aagtcaacat ccctgggttc atccacacct 2340
gtgttttcta tgagttctcc aatatcaagg agatttaaca atttgtttgg taaaacagga 2400
acgtataatg tttccacccc agaagcaacc agctcatccc tggaaaactc atccagtgct 2460
tcttcgttgc tcaactaaga acaggataat ccaacctacg tgacctcccg gggacagtgg 2520
ctgtgctttt aaaaagagat gcttgcaaag caatggggaa cgtgttctcg gggcaggttt 2580
ccgggagcag atgccaaaaa gactttttca tagagaagag gctttctttt gtaaagacag 2640
aataaaaata attgttatgt ttctgtttgt tccctccccc tccccttgt gtgataccac  2700
atgtgtatag tatttaagtg aaactcaagc cctcaaggcc caacttctct gtctatattg 2760
taatatagaa tttcgaagag acattttcac tttttacaca ttgggcacaa agataagctt 2820
tgattaaagt agtaagtaaa aggctaccta ggaaatactt cagtgaattc taagaaggaa 2880
ggaaggaagg aaggagggaa agaagggagg aaaccagga                        2919
```

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 7 :
C12N 15/12, C07K 14/705, C12Q 1/68, A61K 38/17, C07K 16/28, 14/72

A3

(11) International Publication Number: WO 00/15793

(43) International Publication Date: 23 March 2000 (23.03.00)

(21) International Application Number: PCT/US99/20958

(22) International Filing Date: 17 September 1999 (17.09.99)

(30) Priority Data:
09/156,513 17 September 1998 (17.09.98) US

(63) Related by Continuation (CON) or Continuation-in-Part (CIP) to Earlier Application
US 09/156,513 (CIP)
Filed on 17 September 1998 (17.09.98)

(71) Applicant (*for all designated States except US*): INCYTE PHARMACEUTICALS, INC. [US/US]; 3174 Porter Drive, Palo Alto, CA 94304 (US).

(72) Inventors; and
(75) Inventors/Applicants (*for US only*): BANDMAN, Olga [US/US]; 366 Anna Avenue, Mountain View, CA 94043 (US). LAL, Preeti [IN/US]; 2382 Lass Drive, Santa Clara, CA 95054 (US). TANG, Y., Tom [CN/US]; 4230 Ranwick Court, San Jose, CA 95118 (US). CORLEY, Neil, C. [US/US]; 1240 Dale Avenue #30, Mountain View, CA 94040 (US). GUEGLER, Karl, J. [CH/US]; 1048 Oakland Avenue, Menlo Park, CA 94025 (US). GORGONE, Gina, A. [US/US]; 1253 Pinecrest Drive, San Francisco, CA 94132 (US). BAUGHN, Mariah, R. [US/US]; 14244 Santiago Road, San Leandro, CA 94577 (US).

(74) Agents: BILLINGS, Lucy, J. et al.; Incyte Pharmaceuticals, Inc., 3174 Porter Drive, Palo Alto, CA 94304 (US).

(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published
With international search report.

(88) Date of publication of the international search report: 28 September 2000 (28.09.00)

(54) Title: HUMAN GPCR PROTEINS

(57) Abstract

The invention provides human GPCR proteins (HGPRP) and polynucleotides which identify and encode HGPRP. The invention also provides expression vectors, host cells, antibodies, agonists, and antagonists. The invention also provides methods for diagnosing, treating or preventing disorders associated with expression of HGPRP.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

INTERNATIONAL SEARCH REPORT

Internatic Application No
PCT/US 99/20958

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 C12N15/12 C07K14/705 C12Q1/68 A61K38/17 C07K16/28
C07K14/72

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 C07K C12N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	"NCI-CGAP" DATABASE EMBL ID AA877534, ACCESSION NUMBER AA877534, 30 March 1998 (1998-03-30), XP002110947 98,4% identity with seq ID no.7 in 625 bp overlap nt.1184-1810 reverse orientation abstract ---	5,9-12

-/--

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
22 February 2000	28.06.00

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	LE CORNEC N.D.R.

2

Form PCT/ISA/210 (second sheet) (July 1992)

Category *	C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	M.D. ADAMS ET AL: "EST114556 Uterus tumor I Homo sapiens cDNA 5' end" EMBL DATABASE ENTRY HSZZ08200, ACCESSION NUMBER AA303063, 18 April 1997 (1997-04-18), XP002131040 99,% identity in 284 bp overlap : nt616-nt899from seq ID no. 7. abstract & M.D. ADAMS ET AL: "Rapid cDNA sequencing (Expressed sequence tags) from a directionally cloned human infant brain cDNA library" NATURE GENETICS, vol. 4, 1993, pages 373-380, XP000600257 ---	9-13
X	L. HILLIER ET AL: "EST. Homo sapiens cDNA clone 663473 5'" EMBL DATABASE ENTRY HS1143823, ACCESSION NUMBER AA224173, 22 February 1997 (1997-02-22), XP002131041 98,1% identity in 413 bp overlap: nt221-nt628 from seq ID no.7 abstract ---	9-13
A	WO 94 29449 A (SALK INST BIOTECH IND ;DAGGETT LORRIE (US); ELLIS STEVEN B (US); L) 22 December 1994 (1994-12-22) the whole document ---	1-20
A	S. WATSON AND S. ARKINSTAL: THE G-PROTEIN LINKED RECEPTOR FACTSBOOK,1994, XP002131042 GB, London, Academic press cited in the application page 2 -page 6 page 252 -page 260 page 130 -page 141 ---	1-20
A	OLIVEIRA L ET AL: "A COMMON MOTIF IN G-PROTEIN-COUPLED SEVEN TRANSMEMBRANE HELIX RECEPTORS" JOURNAL OF COMPUTER-AIDED MOLECULAR DESIGN,XX,ESCOM SCIENCE PUBLISHERS BV, vol. 7, no. 6, 1 December 1993 (1993-12-01), pages 649-658, XP002050853 ISSN: 0920-654X ---	
A	WO 96 29404 A (SIBIA NEUROSCIENCES INC ;DAGGETT LORRIE P (US); LU CHIN CHUN (US)) 26 September 1996 (1996-09-26) the whole document ---	1-20

-/--

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
P,X	WO 99 06550 A (LACROIX BRUNO ;DUCLERT AYMERIC (FR); GENSET (FR); DUMAS MILNE EDWA) 11 February 1999 (1999-02-11) Seq ID no. 220 is 99% identical to seq ID no.7: nt 842-1047. seq ID no. 498 is 100% identical to seq ID no.1: aa 262-320. claims ---	1-20
E	WO 99 53054 A (SMITHKLINE BEECHAM PLC) 21 October 1999 (1999-10-21) sequences ID 1 and 2 claims ---	1-20
E	WO 99 60121 A (SAVITSKY KINNERET ;COMPUGEN LTD (IL); MINTZ LIAT (IL); TOPORIK AMI) 25 November 1999 (1999-11-25) sequences ID no. 1 and 2 pages 44-49 abstract; claims -----	1-20

International application No.
PCT/US 99/20958

INTERNATIONAL SEARCH REPORT

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. [X] Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

Although claim 19 is directed to a method of treatment of the human/animal body (rule 39.1 IV PCT.), the search has been carried out and based on the alleged effects of the compound/composition.

2. [X] Claims Nos.:
because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:

see FURTHER INFORMATION sheet PCT/ISA/210

3. [ ] Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. [ ] As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.

2. [ ] As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. [ ] As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:

4. [X] No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Claims 1-20, all partially.

Remark on Protest

[ ] The additional search fees were accompanied by the applicant's protest.

[ ] No protest accompanied the payment of additional search fees.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

Continuation of Box I.2

Claims Nos.: 17-18,20 all partially as far as they do not concern the antibody of the polypeptides

Claims 17,18 and 20 refer to an agonist, an antagonist of the polypeptides and a method of treatment or prevention using the antagonist without giving a true technical characterization. Moreover, no such compounds are defined in the application. In consequence, the scope of said claims is ambiguous and vague, and their subject-matter is not sufficiently disclosed and supported.
No search can be carried out for such purely speculative claims whose wording is, in fact, a mere recitation of the results to be achieved.

The applicant's attention is drawn to the fact that claims, or parts of claims, relating to inventions in respect of which no international search report has been established need not be the subject of an international preliminary examination (Rule 66.1(e) PCT). The applicant is advised that the EPO policy when acting as an International Preliminary Examining Authority is normally not to carry out a preliminary examination on matter which has not been searched. This is the case irrespective of whether or not the claims are amended following receipt of the search report or during any Chapter II procedure.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

This International Searching Authority found multiple (groups of) inventions in this international application, as follows:

1. Claims: 1-20 all partially

   A polypeptide comprising the amino acid sequence ID no. 1 encoded by the polynucleotide sequence ID no. 7. Fragments thereof. Method of production by genetic engineering. Host cell. Antibody. Therapeutic uses.

2. Claims: 1-20 all partially

   A polypeptide comprising the amino acid sequence ID no. 2 encoded by the polynucleotide sequence ID no. 8. Fragments thereof. Method of production by genetic engineering. Host cell. Antibody. Therapeutic uses.

3. Claims: 1-20 all partially

   A polypeptide comprising the amino acid sequence ID no. 3 encoded by the polynucleotide sequence ID no. 9. Fragments thereof. Method of production by genetic engineering. Host cell. Antibody. Therapeutic uses.

4. Claims: 1-20 all partially

   A polypeptide comprising the amino acid sequence ID no. 4 encoded by the polynucleotide sequence ID no. 10. Fragments thereof. Method of production by genetic engineering. Host cell. Antibody. Therapeutic uses.

5. Claims: 1-20 all partially

   A polypeptide comprising the amino acid sequence ID no. 6 encoded by the polynucleotide sequence ID no. 12. Fragments thereof. Method of production by genetic engineering. Host cell. Antibody. Therapeutic uses.

INTERNATIONAL SEARCH REPORT

Information on patent family members

Internatio Application No
PCT/US 99/20958

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9429449 A	22-12-1994	US 5521297 A	28-05-1996
		AU 685471 B	22-01-1998
		AU 7098994 A	03-01-1995
		CA 2161811 A	22-12-1994
		EP 0701611 A	20-03-1996
		GB 2286398 A,B	16-08-1995
		JP 8511168 T	26-11-1996
		US 6001581 A	14-12-1999
		US 5912122 A	15-06-1999
		US 5807689 A	15-09-1998
WO 9629404 A	26-09-1996	US 5912122 A	15-06-1999
		AU 5314696 A	08-10-1996
		CA 2215730 A	26-09-1996
		EP 0815227 A	07-01-1998
		JP 11503605 T	30-03-1999
WO 9906550 A	11-02-1999	AU 8555198 A	22-02-1999
		EP 1000148 A	17-05-2000
WO 9953054 A	21-10-1999	NONE	
WO 9960121 A	25-11-1999	AU 3952699 A	06-12-1999

```
                                                   Aat II
TGGCTGACCGCCCAACGACCCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCC 3600
ACCGACTGGCGGGTTGCTGGGGGCGGGTAACTGCAGTTATTACTGCATACAAGGGTATCATTGCGGTTATCCCTGAAAGG

L A D R P T T P A H . R Q . . R M F P . . R Q . G L S
W L T A Q R P P P I D V N N D V C S H S N A N R D F P
G . P P N D P R P L T S I M T Y V P I V T P I G T F

A S R G V V G A W Q R . Y H R I N G Y Y R W Y P S E
Q S V A W R G G G M S T L L S T H E W L L A L L S K G
P Q G G L S G R G N V D I I V Y T G M T V G I P V K W

   Aat II                                             Bgl I  Rsa I           Nde I    Rsa I
ATTGACGTCAATGGGTGGACTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCC 3680
TAACTGCAGTTACCCACCTGATAAATGCCATTTGACGGGTGAACCGTCATGTAGTTCACATAGTATACGGTTCATGCGGG

I D V N G W T I Y G K L P T W Q Y I K C I I C Q V R P
L T S M G G L F T V N C P L G S T S S V S Y A K Y A
H . R Q W V D Y L R . T A H L A V H Q V Y H M P S T P

M S T L P H V I . P L S G V Q C Y M L H I M H W T R G
N V D I P P S N V T F Q G S P L V D L T D Y A L Y A G
Q R . H T S . K R Y V A W K A T C . T Y . I G L V G

      Aat II                Hae III Bgl I           Rsa I
CCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTATGGGACTTTCCTACTTGGCA 3760
GGATAACTGCAGTTACTGCCATTTACCGGGCGGACCGTAATACGGGTCATGTACTGGAATACCCTGAAAGGATGAACCGT

L L T S M T V N G P P G I M P S T . P Y G T F L L G
P Y . R Q . R . M A R L A L C P V H D L M G L S Y L A
P I D V N D G K W P A W H Y A Q Y M T L W D F P T W Q

R N V D I V T F P G G P M I G L V H G . P V K R S P L
. Q R . H R Y I A R R A N H G T C S R I P S E . K A
G I S T L S P L H G A Q C . A W Y M V K H S K G V Q C

         BsaA I               Nco I
Rsa I    SnaB I               Sty I                     Rsa I
GTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACATCAATGGGCGTGGATAGCGGTTTG 3840
CATGTAGATGCATAATCAGTAGCGATAATGGTACCACTACGCCAAAACCGTCATGTAGTTACCCGCACCTATCGCCAAAC

S T S T Y . S S L L P W . C G F G S T S M G V D S G L
V H L R I S H R Y Y H G D A V L A V H Q W A W I A V .
Y I Y V L V I A I T M V M R F W Q Y I N G R G . R F

V D V Y . D D S N G H H H P K P L V D I P T S L P K
T C R R I L . R . . W P S A T K A T C . H A H I A T Q
Y M . T N T M A I V M T I R N Q C Y M L P R P Y R N S
```

FIGURE 3K

```
                                  ApaL I
ACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTT
-------------------------------------------------------------------------------- 3280
TGGCGACAACTCTAGGTCAAGCTACATTGGGTGAGCACGTGGGTTGACTAGAAGTCGTAGAAAATGAAAGTGGTCGCAAA

  P  L  L  R  S  S  S  M  .  P  T  R  A  P  N  .  S  S  A  S  F  T  F  T  S  V
 Y  R  C  .  D  P  V  R  C  N  P  L  V  H  P  T  D  L  Q  H  L  L  L  S  P  A  F
T  A  V  E  I  Q  F  D  V  T  H  S  C  T  Q  L  I  F  S  I  F  Y  F  H  Q  R  F
--------------------------------------------------------------------------------
  G  S  N  L  D  L  E  I  Y  G  V  R  A  G  L  Q  D  E  A  D  K  V  K  V  L  T  E
 R  Q  Q  S  G  T  R  H  L  G  S  T  C  G  V  S  R  .  C  R  K  S  E  G  A  N
V  A  T  S  I  W  N  S  T  V  W  E  H  V  W  S  I  K  L  M  K  .  K  .  W  R  K

CTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTGAATACTCATACTC
-------------------------------------------------------------------------------- 3360
GACCCACTCGTTTTTGTCCTTCCGTTTTACGGCGTTTTTTCCCTTATTCCCGCTGTGCCTTTACAACTTATGAGTATGAG

S  G  .  A  K  T  G  R  Q  N  A  A  K  K  G  I  R  A  T  R  K  C  .  I  L  I  L
 L  G  E  Q  K  Q  E  G  K  M  P  Q  K  R  E  .  G  R  H  G  N  V  E  Y  S  Y  S
  W  V  S  K  N  R  K  A  K  C  R  K  K  G  N  K  G  D  T  E  M  L  N  T  H  T
--------------------------------------------------------------------------------
  P  H  A  F  V  P  L  C  F  A  A  F  F  P  I  L  A  V  R  F  H  Q  I  S  M  S
R  P  S  C  F  C  S  P  L  I  G  C  F  L  S  Y  P  R  C  P  F  T  S  Y  E  Y  E
 Q  T  L  L  F  L  F  A  F  H  R  L  F  P  F  L  P  S  V  S  I  N  F  V  .  V  R

                                                      Hinc II    Spe I    AseI
TTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGCGCGTTGACATTGATTATTGACTAGTTATTAA
-------------------------------------------------------------------------------- 3440
AAGGAAAAAGTTATAATAACTTCGTAAATAGTCCCAATAACAGAGTACGCGCAACTGTAACTAATAACTGATCAATAATT

 F  L  F  Q  Y  Y  .  S  I  Y  Q  G  Y  C  L  M  R  V  D  I  D  Y  .  L  V  I  N
  S  F  F  N  I  I  E  A  F  I  R  V  I  V  S  C  A  L  T  L  I  I  D  .  L  L
L  P  F  S  I  L  L  K  H  L  S  G  L  L  S  H  A  R  .  H  .  L  L  T  S  Y  .
--------------------------------------------------------------------------------
K  R  K  .  Y  .  Q  L  M  .  .  P  .  Q  R  M  R  T  S  M  S  .  Q  S  T  I  L
 E  K  K  L  I  I  S  A  N  I  L  T  I  T  E  H  A  N  V  N  I  I  S  .  N  N  I
  G  K  E  I  N  N  F  C  K  D  P  N  N  D  .  A  R  Q  C  Q  N  N  V  L  .  .

                                                                 Hae III  Bgl I
TAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATATATGGAGTTCCGCGTTACATAACTTACGGTAAATGGCCCGCC
-------------------------------------------------------------------------------- 3520
ATCATTAGTTAATGCCCCAGTAATCAAGTATCGGGTATATACCTCAAGGCGCAATGTATTGAATGCCATTTACCGGGCGG

  S  N  Q  L  R  G  H  .  F  I  A  H  I  W  S  S  A  L  H  N  L  R  .  M  A  R
I  V  I  N  Y  G  V  I  S  S  .  P  I  Y  G  V  P  R  Y  I  T  Y  G  K  W  P  A
 .  .  S  I  T  G  S  L  V  H  S  P  Y  M  E  F  R  V  T  .  L  T  V  N  G  P  P
--------------------------------------------------------------------------------
L  L  .  N  R  P  .  .  N  M  A  W  I  H  L  E  A  N  C  L  K  R  Y  I  A  R  R
  T  I  L  .  P  T  M  L  E  Y  G  M  Y  P  T  G  R  .  M  V  .  P  L  H  G  A
Y  Y  D  I  V  P  D  N  T  .  L  G  Y  I  S  N  R  T  V  Y  S  V  T  F  P  G  G
```

FIGURE 3J

```
                                                                          Ava II           Pvu I                    Hae III
CGAGTTACATGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGC 2960
GCTCAATGTACTAGGGGGTACAACACGTTTTTTCGCCAATCGAGGAAGCCAGGAGGCTAGCAACAGTCTTCATTCAACCG

R V T . S P M L C K K A V S S F G P P I V V R S K L A
E L H D P P C C A K K R L A P S V L R S L S E V S W
A S Y M I P H V V Q K S G . L L R S S D R C Q K . V G

R T V H D G M N H L F A T L E K P G G I T T L L L N A
S N C S G G H Q A F F R N A G E T R R D N D S T L Q G
L . M I G W T T C F L P . S R R D E S R Q . F Y T P

CGCAGTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTCATGCCATCCGTAAGATGCTTTTCTGTGA 3040
GCGTCACAATAGTGAGTACCAATACCGTCGTGACGTATTAAGAGAATGACAGTACGGTAGGCATTCTACGAAAAGACACT

A V L S L M V M A A L H N S L T V M P S V R C F S V
P Q C Y H S W L W Q H C I I L L L S C H P . D A F L .
R S V I T H G Y G S T A . F S Y C H A I R K M L F C D

A T N D S M T I A A S C L E R V T M G D T L H K E T V
C H . . E H N H C C Q M I R K S D H W G Y S A K R H
R L T I V . P . P L V A Y N E . Q . A M R L I S K Q S

       Rsa I
       Sca I                                                        Nci I      Hinc II
CTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAACACGGGAT 3120
GACCACTCATGAGTTGGTTCAGTAAGACTCTTATCACATACGCCGCTGGCTCAACGAGAACGGGCCGCAGTTGTGCCCTA

T G E Y S T K S F . E . C M R R P S C S C P A S T R D
L V S T Q P S H S E N S V C G D R V A L A R R Q H G I
W . V L N Q V I L R I V Y A A T E L L L P G V N T G

P S Y E V L D N Q S Y H I R R G L Q E Q G A D V R S
S T L V . G L . E S F L T H P S R T A R A R R . C P I
Q H T S L W T M R L I T Y A A V S N S K G P T L V P Y

                           Dra I                   Xmn I
AATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTT 3200
TTATGGCGCGGTGTATCGTCTTGAAATTTTCACGAGTAGTAACCTTTTGCAAGAAGCCCCGCTTTTGAGAGTTCCTAGAA

N T A P H S R T L K V L I I G K R S S G R K L S R I L
I P R H I A E L . K C S S L E N V L R G E N S Q G S
. Y R A T . Q N F K S A H H W K T F F G A K T L K D L

L V A G C L L V K F T S M M P F R E E P R F S E L I K
I G R W M A S S . F H E D N S F T R R P S F E . P D .
Y R A V Y C F K L L A . . Q F V N K P A F V R L S R
```

FIGURE 3I

```
AACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCGTTCATCCATAGTTG
+--------+---------+---------+---------+---------+---------+---------+---------+ 2560
TTGAACCAGACTGTCAATGGTTACGAATTAGTCACTCCGTGGATAGAGTCGCTAGACAGATAAAGCAAGTAGGTATCAAC

 T  W  S  D  S  Y  Q  C  L  I  S  E  A  P  I  S  A  I  C  L  F  R  S  S  I  V
K  L  G  L  T  V  T  N  A  .  S  V  R  H  L  S  Q  R  S  V  Y  F  V  H  P  .  L
 N  L  V  .  Q  L  P  M  L  N  Q  .  G  T  Y  L  S  D  L  S  I  S  F  I  H  S  C
+--------+---------+---------+---------+---------+---------+---------+---------+
 V  Q  D  S  L  .  W  H  K  I  L  S  A  G  I  E  A  I  Q  R  N  R  E  D  M  T  A
  S  P  R  V  T  V  L  A  .  D  T  L  C  R  D  .  R  D  T  .  K  T  .  G  Y  N
F  K  T  Q  C  N  G  I  S  L  .  H  P  V  .  R  L  S  R  D  I  E  N  M  W  L  Q

                                                Hae III
CCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGAC
+--------+---------+---------+---------+---------+---------+---------+---------+ 2640
GGACTGAGGGGCAGCACATCTATTGATGCTATGCCCTCCCGAATGGTAGACCGGGGTCACGACGTTACTATGGCGCTCTG

A  .  L  P  V  V  .  I  T  T  I  R  E  G  L  P  S  G  P  S  A  A  M  I  P  R  D
 P  D  S  P  S  C  R  .  L  R  Y  G  R  A  Y  H  L  A  P  V  L  Q  .  Y  R  E  T
  L  T  P  R  R  V  D  N  Y  D  T  G  G  L  T  I  W  P  Q  C  C  N  D  T  A  R
+--------+---------+---------+---------+---------+---------+---------+---------+
 Q  S  G  T  T  Y  I  V  V  I  R  S  P  K  G  D  P  G  L  A  A  I  I  G  R  S
G  S  E  G  D  H  L  Y  S  R  Y  P  L  A  .  W  R  A  G  T  S  C  H  Y  R  S  V
 R  V  G  R  R  T  S  L  .  S  V  P  P  S  V  M  Q  G  W  H  Q  L  S  V  A  L  G

                                              Bgl I
                                                 Hae III           Ava II
CCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTT
+--------+---------+---------+---------+---------+---------+---------+---------+ 2720
GGTGCGAGTGGCCGAGGTCTAAATAGTCGTTATTTGGTCGGTCGGCCTTCCCGGCTCGCGTCTTCACCAGGACGTTGAAA

 P  R  S  P  A  P  D  L  S  A  I  N  Q  P  A  G  R  A  E  R  R  S  G  P  A  T  L
  H  A  H  R  L  Q  I  Y  Q  Q  .  T  S  Q  P  E  G  P  S  A  E  V  V  L  Q  L
P  T  L  T  G  S  R  F  I  S  N  K  P  A  S  R  K  G  R  A  Q  K  W  S  C  N  F
+--------+---------+---------+---------+---------+---------+---------+---------+
G  R  E  G  A  G  S  K  D  A  I  F  W  G  A  P  L  A  S  R  L  L  P  G  A  V  K
 W  A  .  R  S  W  I  .  .  C  Y  V  L  W  G  S  P  G  L  A  S  T  T  R  C  S  .
  V  S  V  P  E  L  N  I  L  L  L  G  A  L  R  F  P  R  A  C  F  H  D  Q  L  K

                  AseI      Nci I                                    Fsp I
ATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTG
+--------+---------+---------+---------+---------+---------+---------+---------+ 2800
TAGGCGGAGGTAGGTCAGATAATTAACAACGGCCCTTCGATCTCATTCATCAAGCGGTCAATTATCAAACGCGTTGCAAC

 S  A  S  I  Q  S  I  N  C  C  R  E  A  R  V  S  S  S  P  V  N  S  L  R  N  V
Y  P  P  P  S  S  L  L  I  V  A  G  K  L  E  .  V  V  R  Q  L  I  V  C  A  T  L
 I  R  L  H  P  V  Y  .  L  L  P  G  S  .  S  K  .  F  A  S  .  .  F  A  Q  R  C
+--------+---------+---------+---------+---------+---------+---------+---------+
 D  A  E  M  W  D  I  L  Q  Q  R  S  A  L  T  L  L  E  G  T  L  L  K  R  L  T  T
  G  G  G  D  L  R  N  I  T  A  P  F  S  S  Y  T  T  R  W  N  I  T  Q  A  V  N
I  R  R  W  G  T  .  .  N  N  G  P  L  .  L  L  Y  N  A  L  .  Y  N  A  C  R  Q

TGCCATTGCTACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTTCCCAACGATCAAGG
+--------+---------+---------+---------+---------+---------+---------+---------+ 2880
ACGGTAACGATGTCCGTAGCACCACAGTGCGAGCAGCAAACCATACCGAAGTAAGTCGAGGCCAAGGGTTGCTAGTTCC

 A  I  A  T  G  I  V  V  S  R  S  S  F  G  M  A  S  F  S  S  G  S  Q  R  S  R
L  P  L  L  Q  A  S  W  C  H  A  R  R  L  V  W  L  H  S  A  P  V  P  N  D  Q  G
 C  H  C  Y  R  H  R  G  V  T  L  V  V  W  Y  G  F  I  Q  L  R  F  P  T  I  K
+--------+---------+---------+---------+---------+---------+---------+---------+
A  M  A  V  P  M  T  T  D  R  E  D  N  P  I  A  E  N  L  E  P  E  W  R  D  L
 G  N  S  C  A  D  H  H  .  A  R  R  K  T  H  S  .  E  A  G  T  G  L  S  .  P
Q  W  Q  .  L  C  R  P  T  V  S  T  T  Q  [illegible]  R  N  G  V  I  L  A
```

FIGURE 3H

```
                                                                                Hae III
TGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAAC 2160
ACCGTCGTCGGTGACCATTGTCCTAATCGTCTCGCTCCATACATCCGCCACGATGTCTCAAGAACTTCACCACCGGATTG

L A A A T G N R I S R A R Y V G G A T E F L K W W P N
W Q Q P L V T G L A E R G M . A V L Q S S . S G G L T
G S S H W . Q D . Q S E V C R R C Y R V L E V V A .

A A A V P L L I L L A L Y T P P A V S N K F H H G L
Q C C G S T V P N A S R P I Y A T S C L E Q L P P R V
P L L W Q Y C S . C L S T H L R H . L T R S T T A . S

TACGGCTACACTAGAAGGACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTC 2240
ATGCCGATGTGATCTTCCTGTCATAAACCATAGACGCGAGACGACTTCGGTCAATGGAAGCCTTTTTCTCAACCATCGAG

Y G Y T R R T V F G I C A L L K P V T F G K R V G S S
T A T L E G Q Y L V S A L C . S Q L P S E K E L V A
L R L H . K D S I W Y L R S A E A S Y L R K K S W . L

. P . V L L V T N P I Q A R S F G T V K P F L T P L E
V A V S S P C Y K T D A S Q Q L W N G E S F S N T A R
R S C . F S L I Q Y R R E A S A L . R R F F L Q Y S

TTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGAT 2320
AACTAGGCCGTTTGTTTGGTGGCGACCATCGCCACCAAAAAAACAAACGTTCGTCGTCTAATGCGCGTCTTTTTTTCCTA

. S G K Q T T A G S G G F F V C K Q Q I T R R K K G
L D P A N K P P L V A V V F L F A S S R L R A E K K D
L I R Q T N H R W . R W F F C L Q A A D Y A Q K K R I

Q D P L C V V A P L P P K K T Q L C C I V R L F F P D
S G A F L G G S T A T T K K N A L L L N R A S F F S
K I R C V F W R Q Y R H N K Q K C A A S . A C F F L I

                                                                                  BspH I
CTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATG 2400
GAGTTCTTCTAGGAAACTAGAAAAGATGCCCCAGACTGCGAGTCACCTTGCTTTTGAGTGCAATTCCCTAAAACCAGTAC

S Q E D P L I F S T G S D A Q W N E N S R . G I L V M
L K K I L . S F L R G L T L S G T K T H V K G F W S .
S R R S F D L F Y G V . R S V E R K L T L R D F G H

. S S G K I K E V P D S A . H F S F E R . P I K T M
R L F I R Q D K R R P R V S L P V F V . T L P N Q D H
E L L D K S R K . P T Q R E T S R F S V N L S K P . S

                                               Dra I                Dra I
AGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTA 2480
TCTAATAGTTTTTCCTAGAAGTGGATCTAGGAAAATTTAATTTTTACTTCAAAATTTAGTTAGATTTCATATATACTCAT

R L S K R I F T . I L L N . K . S F K S I . S I Y E .
D Y Q K G S S S P R S F . I K N E V L N Q S K V Y M S
E I I K K D L H L D P F K L K M K F . I N L K Y I . V

L N D F L I K V . I R K F . F H L K L D I . L I Y S Y
S . . F P D E G L D K . I L F S T K F . D L T Y I L L
I I L F S R . R S G K L N F I F N . I L R F Y I H T
```

FIGURE 3G

Hae III Hae III Hae III

TGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGA 1760

ACTCGTTTTCCGGTCGTTTTCCGGTCCTTGGCATTTTTCCGGCGCAACGACCGCAAAAAGGTATCCGAGGCGGGGGGACT

. A K G Q Q K A R N R K K A A L L A F F H R L R P P D

E Q K A S K R P G T V K R P R C W R F S I G S A P L

V S K R P A K G Q E P . K G R V A G V F P . A P P P .

H A F P W C F A L F R L F A A N S A N K W L S R G G S

S C F A L L L G P V T F L G R Q Q R K E M P E A G R V

L L L G A F P W S G Y F P R T A P T K G Y A G G G Q

CGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTG 1840

GCTCGTAGTGTTTTTAGCTGCGAGTTCAGTCTCCACCGCTTTGGGCTGTCCTGATATTTCTATGGTCCGCAAAGGGGGAC

E H H K N R R S S Q R W R N P T G L . R Y Q A F P P

T S I T K I D A Q V R G G E T R Q D Y K D T R R F P L

R A S Q K S T L K S E V A K P D R T I K I P G V S P W

S C . L F R R E L . L H R F G V P S Y L Y W A N G G P

L M V F I S A . T L P P S V R C S . L S V L R K G R

R A D C F D V S L D S T A F G S L V I F I G P T E G Q

GAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTG 1920

CTTCGAGGGAGCACGCGAGAGGACAAGGCTGGGACGGCGAATGGCCTATGGACAGGCGGAAAGAGGGAAGCCCTTCGCAC

G S S L V R S P V P T L P L T G Y L S A F L P S G S V

E A P S C A L L F R P C R L P D T C P P F S L R E A W

K L P R A L S C S D P A A Y R I P V R L S P F G K R

L E R T R E G T G V R G S V P Y R D A K R G E P L T

S A G E H A R R N R G Q R K G S V Q G G K E R R S A H

F S G R A S E Q E S G A A . R I G T R R E G K P F R P .

ApaL I

GCGCTTTCTCAATGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACC 2000

CGCGAAAGAGTTACGAGTGCGACATCCATAGAGTCAAGCCACATCCAGCAAGCGAGGTTCGACCCGACACACGTGCTTGG

A L S Q C S R C R Y L S S V . V V R S K L G C V H E P

R F L N A H A V G I S V R C R S F A P S W A V C T N

G A F S M L T L . V S Q F G V G R S L Q A G L C A R T

A S E . H E R Q L Y R L E T Y T T R E L S P Q T C S G

R K R L A . A T P I E T R H L D N A G L Q A T H V F G

A K E I S V S Y T D . N P T P R E S W A P S H A R V

Nci I

CCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCAC 2080

GGGGCAAGTCGGGCTGGCGACGCGGAATAGGCCATTGATAGCAGAACTCAGGTTGGGCCATTCTGTGCTGAATAGCGGTG

P V Q P D R C A L S G N Y R L E S N P V R H D L S P

P P F S P T A A P Y P V T I V L S P T R . D T T Y R H

P R S A R P L R L I R . L S S . V Q P G K T R L I A T

G T . G S R Q A K D P L . R R S D L G T L C S K D G S

G N L G V A A G . G T V I T K L G V R Y S V V . R W

G R E A R G S R R I R Y S D D Q T W G P L V R S I A V

FIGURE 3F

```
  BsrB I
GTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCTAATGAGTGAGCTAA 1440
CAATAGGCGAGTGTTAAGGTGTGTTGTATGCTCGGCCTTCGTATTTCACATTTCGGACCCCACGGATTACTCACTCGATT

C Y P L T I P H N I R A G S I K C K A W G A . . V S .
V I R S Q F H T T Y E P E A . S V K P G V P N E . A N
L S A H N S T Q H T S R K H K V . S L G C L M S E L

. G S V I G C L M R A P L M F H L A Q P A . H T L .
T I R E C N W V V Y S G S A Y L T F G P T G L S H A L
N D A . L E V C C V L R F C L T Y L R P H R I L S S V

  AseI                                                      Pvu II AseI     Hae III
CTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCA 1520
GAGTGTAATTAACGCAACGCGAGTGACGGGCGAAAGGTCAGCCCTTTGGACAGCACGGTCGACGTAATTACTTAGCCGGT

L T L I A L R S L P A F Q S G N L S C Q L H . . I G Q
S H . L R C A H C P L S S R E T C R A S C I N E S A
T H I N C V A L T A R F P V G K P V V P A A L M N R P

S V N I A N R E S G A K W D P F R D H W S C . H I P W
E C . . N R Q A . Q G S E L R S V Q R A L Q M L S D A L
. M L Q T A S V A R K G T P F G T T G A A N I F R G

                                             Sap I
ACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGC 1600
TGCGCGCCCCTCTCCGCCAAACGCATAACCCGCGAGAAGGCGAAGGAGCGAGTGACTGAGCGACGCGAGCCAGCAAGCCG

R A G R G G L R I G R S S A S S L T D S L R S V V R
N A R G E A V C V L G A L P L P R S L T R C A R S F G
T R G E R R F A Y W A L F R F L A H . L A A L G R S A

R A P L P P K R I P R E E A E E S V S E S R E T T R S
A R P S A T Q T N P A R G S G R E S V R Q A R D N P
V R P S L R N A Y Q A S K R K R A . Q S A A S P R E A

  BsrB I
TGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACATG 1680
ACGCCGCTCGCCATAGTCGAGTGAGTTTCCGCCATTATGCCAATAGGTGTCTTAGTCCCCTATTGCGTCCTTTCTTGTAC

L R R A V S A H S K A V I R L S T E S G D N A G K N M
C G E R Y Q L T Q R R . Y G Y P Q N Q G I T Q E R T C
A A S G I S S L K G G N T V I H R I R G . R R K E H

R R A T D A . E F A T I R N D V S D P S L A P F F M
Q P S R Y . S V . L R Y Y P . G C F . P I V C S L V H
A A L P I L E S L P P L V T I W L I L P Y R L F S C T
```

FIGURE 3E

```
                                                                          Sph I
                                                                           Nsi I
CTCAATTAGTCAGCAACCAGGTGTGGAAAGTCCCCAGGCTCCCCAGCAGGCAGAAGTATGCAAAGCATGCATCTCAATTA
---------------------------------------------------------------------------------- 1120
GAGTTAATCAGTCGTTGGTCCACACCTTTCAGGGGTCCGAGGGGTCGTCCGTCTTCATACGTTTCGTACGTAGAGTTAAT

 S  I  S  Q  Q  P  G  V  E  S  P  Q  A  P  Q  Q  A  E  V  C  K  A  C  I  S  I
S  Q  L  V  S  N  Q  V  W  K  V  P  R  L  P  S  R  Q  K  Y  A  K  H  A  S  Q  L
 L  N  .  S  A  T  R  C  G  K  S  P  G  S  P  A  G  R  S  M  Q  S  M  H  L  N  .
----------------------------------------------------------------------------------
 E  I  L  .  C  G  P  T  S  L  G  W  A  G  W  C  A  S  T  H  L  A  H  M  E  I  L
 .  N  T  L  L  W  T  H  F  T  G  L  S  G  L  L  C  F  Y  A  F  C  A  D  .  N
R  L  .  D  A  V  L  H  P  F  D  G  P  E  G  A  P  L  L  I  C  L  M  C  R  L  .

                                                                              Nco I
                                                                              Sty I
GTCAGCAACCATAGTCCCGCCCCTAACTCCGCCCATCCCGCCCCTAACTCCGCCCAGTTCCGCCCATTCTCCGCCCCATG
---------------------------------------------------------------------------------- 1200
CAGTCGTTGGTATCAGGGCGGGGATTGAGGCGGGTAGGGCGGGGATTGAGGCGGGTCAAGGCGGGTAAGAGGCGGGGTAC

S  Q  Q  P  .  S  R  P  .  L  R  P  S  R  P  .  L  R  P  V  P  P  I  L  R  P  M
 V  S  N  H  S  P  A  P  N  S  A  H  P  A  P  N  S  A  Q  F  R  P  F  S  A  P  W
  S  A  T  I  V  P  P  L  T  P  P  I  P  P  L  T  P  P  S  S  A  H  S  P  P  H
----------------------------------------------------------------------------------
  .  C  G  Y  D  R  G  .  S  R  G  D  R  G  .  S  R  G  T  G  G  M  R  R  G  M
T  L  L  W  L  G  A  G  L  E  A  W  G  A  G  L  E  A  W  N  R  G  N  E  A  G  H
 D  A  V  M  T  G  G  R  V  G  G  M  G  G  R  V  G  G  L  E  A  W  E  G  G  W  P

                                                 Bgl I
                                Hae III Hae III       Hae III
GCTGACTAATTTTTTTTATTTATGCAGAGGCCGAGGCCGCCTCGGCCTCTGAGCTATTCCAGAAGTAGTGAGGAGGCTTT
---------------------------------------------------------------------------------- 1280
CGACTGATTAAAAAAAATAAATACGTCTCCGGCTCCGGCGGAGCCGGAGACTCGATAAGGTCTTCATCACTCCTCCGAAA

 A  D  .  F  F  L  F  M  Q  R  P  R  P  P  R  P  L  S  Y  S  R  S  S  E  E  A  F
  L  T  N  F  F  Y  L  C  R  G  R  G  R  L  G  L  .  A  I  P  E  V  V  R  R  L
G  .  L  I  F  F  I  Y  A  E  A  E  A  A  S  A  S  E  L  F  Q  K  .  .  G  G  F
----------------------------------------------------------------------------------
A  S  .  N  K  K  N  I  C  L  G  L  G  G  R  G  R  L  .  E  L  L  L  S  S  A  K
 S  V  L  K  K  .  K  H  L  P  R  P  R  R  P  R  Q  A  I  G  S  T  T  L  L  S  K
  Q  S  I  K  K  I  .  A  S  A  S  A  A  E  A  E  S  S  N  W  F  Y  H  P  P  K

     Hae III
     Stu I
      Avr II
      Sty I                      Ava I
                                 Xho I
TTTGGAGGCCTAGGCTTTTGCAAAAAGCTCCCTCGAGAGCTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAATT
---------------------------------------------------------------------------------- 1360
AAACCTCCGGATCCGAAAACGTTTTTCGAGGGAGCTCTCGAACCGCATTAGTACCAGTATCGACAAAGGACACACTTTAA

  L  E  A  .  A  F  A  K  S  S  L  E  S  L  A  .  S  W  S  .  L  F  P  V  .  N
F  W  R  P  R  L  L  Q  K  A  P  S  R  A  W  R  N  H  G  H  S  C  F  L  C  E  I
 F  G  G  L  G  F  C  K  K  L  P  R  E  L  G  V  I  M  V  I  A  V  S  C  V  K  L
----------------------------------------------------------------------------------
K  S  A  .  A  K  A  F  L  E  R  S  L  K  A  Y  D  H  D  Y  S  N  G  T  H  F  Q
 Q  L  G  L  S  K  C  F  A  G  E  L  A  Q  R  L  .  P  .  L  Q  K  R  H  S  I
K  P  P  R  P  K  Q  L  F  S  G  R  S  S  P  T  I  M  T  M  A  T  E  Q  T  F  N
```

FIGURE 3D

```
     BsaB I                                        Dra I
TAGAGATCATAATCAGCCATACCACATTTGTAGAGGTTTTACTTGCTTTAAAAAACCTCCCACACCTCCCCCTGAACCTG 800
ATCTCTAGTATTAGTCGGTATGGTGTAAACATCTCCAAAATGAACGAAATTTTTTGGAGGGTGTGGAGGGGGACTTGGAC

 . R S . S A I P H L . R F Y L L . K T S H T S P . T .
  R D H N Q P Y H I C R G F T C F K K P P T P P P E P
L E I I I S H T T F V E V L L A L K N L P H L P L N L

 . L D Y D A M G C K Y L N . K S . F V E W V E G Q V Q
  L S . L . G Y W M Q L P K V Q K L F G G V G G G S G S
  S I M I L W V V N T S T K S A K F F R G C R G R F R

                                 Hinc II
                    MfeI         Hpa I
AAACATAAAATGAATGCAATTGTTGTTGTTAACTTGTTTATTGCAGCTTATAATGGTTACAAATAAAGCAATAGCATCAC 880
TTTGTATTTTACTTACGTTAACAACAACAATTGAACAAATAACGTCGAATATTACCAATGTTTATTTCGTTATCGTAGTG

  N I K . M Q L L L L T C L L Q L I M V T N K A I A S
E T . N E C N C C C . L V Y C S L . W L Q I K Q . H H
 K H K M N A I V V V N L F I A A Y N G Y K . S N S I T

 F M F H I C N N N N V Q K N C S I I T V F L A I A D C
  V Y F S H L Q Q Q . S T . Q L K Y H N C I F C Y C .
F C L I F A I T T T L K N I A A . L P . L Y L L L M V

                                                                          Xba I
AAATTTCACAAATAAAGCATTTTTTTCACTGCATTCTAGTTGTGGTTTGTCCAAACTCATCAATGTATCTTATCATGTCT 960
TTTAAAGTGTTTATTTCGTAAAAAAAGTGACGTAAGATCAACACCAAACAGGTTTGAGTAGTTACATAGAATAGTACAGA

Q I S Q I K H F F H C I L V V V C P N S S M Y L I M S
 K F H K . S I F F T A F . L W F V Q T H Q C I L S C L
  N F T N K A F F S L H S S C G L S K L I N V S Y H V

  I E C I F C K K . Q M R T T T Q G F E D I Y R I M D
L N . L Y L M K K V A N . N H N T W V . . H I K D H R
 F K V F L A N K E S C E L Q P K D L S M L T D . . T .

                                                                           Sph I
Bgl II                                                                       Nsi I
AGATCTTGTGGAATGTGTGTCAGTTAGGGTGTGGAAAGTCCCCAGGCTCCCCAGCAGGCAGAAGTATGCAAAGCATGCAT 1040
TCTAGAACACCTTACACACAGTCAATCCCACACCTTTCAGGGGTCCGAGGGGTCGTCCGTCTTCATACGTTTCGTACGTA

 R S C G M C V S . G V E S P Q A P Q Q A E V C K A C I
  D L V E C V S V R V W K V P R L P S R Q K Y A K H A
. I L W N V C Q L G C G K S P G S P A G R S H Q S M H

L D Q P I H T L . P T S L G W A G W C A S T H L A H M
 S R T S H T D T L T H F T G L S G L L C F Y A F C A D
 I K H F T H . N P H P F D G P E G A P L L I C L M C
```

FIGURE 3C

```
TTTGCTCAGAAGAAATGCCATCTAGTGATGATGAGGCTACTGCTGACTCTCAACATTCTACTCCTCCAAAAAAGAAGAGA 400
AAACGAGTCTTCTTTACGGTAGATCACTACTACTCCGATGACGACTGAGAGTTGTAAGATGAGGAGGTTTTTTCTTCTCT

L L R R N A I . . . . G Y C . L S T F Y S S K K E E
F C S E E M P S S D D E A T A D S Q H S T P P K K K R
F A Q K K C H L V M M R L L L T L N I L L L Q K R R E
K S L L F A M . H H H P . Q Q S E V N . E E L F S S F
Q E S S I G D L S S S A V A S E . C E V G G F F F L
K A . F F H W R T I I L S S S V R L M R S R W F L L S

Sty I
AAGGTAGAAGACCCCAAGGACTTTCCTTCAGAATTGCTAAGTTTTTTGAGTCATGCTGTGTTTAGTAATAGAACTCTTGC 480
TTCCATCTTCTGGGGTTCCTGAAAGGAAGTCTTAACGATTCAAAAAACTCAGTACGACACAAATCATTATCTTGAGAACG

K G R R P Q G L S F R I A K F F E S C C V . . . N S C
K V E D P K D F P S E L L S F L S H A V F S N R T L A
R . K T P R T F L Q N C . V F . V M L C L V I E L L
P L L G W P S E K L I A L N K S D H Q T . Y Y F E Q
F T S S G L S K G E S N S L K K L . A T N L L L V R A
L Y F V G L V K R . F Q . T K Q T M S H K T I S S K S

TTGCTTTGCTATTTACACCACAAAGGAAAAAGCTGCACTGCTATACAAGAAAATTATGGAAAAATATTCTGTAACCTTTA 560
AACGAAACGATAAATGTGGTGTTTCCTTTTTCGACGTGACGATATGTTCTTTTAATACCTTTTTATAAGACATTGGAAAT

L L C Y L H H K G K S C T A I Q E N Y G K I F C N L Y
C F A I Y T T K E K A A L L Y K K I M E K Y S V T F
L A L L F T P Q R K K L H C Y T R K L W K N I L . P L
K S Q . K C W L P F L Q V A I C S F . P F I N Q L R .
Q K A I . V V F S F A A S S Y L F I I S F Y E T V K I
A K S N V G C L F F S C Q . V L F N H F F I R Y G K

AseI
TAAGTAGGCATAACAGTTATAATCATAACATACTGTTTTTTCTTACTCCACACAGGCATAGAGTGTCTGCTATTAATAAC 640
ATTCATCCGTATTGTCAATATTAGTATTGTATGACAAAAAAGAATGAGGTGTGTCCGTATCTCACAGACGATAATTATTG

K . A . Q L . S . H T V F S Y S T Q A . S V C Y . .
I S R H N S Y N H N I L F F L T P H R H R V S A I N N
. V G I T V I I I T Y C F F L L H T G I E C L L L I T
L Y A Y C N Y D Y C V T K E . . E V C A Y L T Q . . Y S
L L C L L . L . L M S N K R V G C L C L T D A I L L
Y T P M V T I I M V Y Q K K K S W V P M S H R S N I V

Rsa I
TATGCTCAAAAATTGTGTACCTTTAGCTTTTTAATTTGTAAAGGGGTTAATAAGGAATATTTGATGTATAGTGCCTTGAC 720
ATACGAGTTTTTAACACATGGAAATCGAAAAATTAAACATTTCCCCAATTATTCCTTATAAACTACATATCACGGAACTG

L C S K I V Y L . L F N L . R G . . G I F D V . C L D
Y A Q K L C T F S F L I C K G V N K E Y L M Y S A L T
M L K N C V P L A F . F V K G L I R N I . C I V P .
H E F I T Y R . S K L K Y L P . Y P I N S T Y H R S
. A . F N H V K L K K I Q L P T L L S Y K I Y L A K V
I S L F Q T G K A K . N T F P N I L F I Q H I T G Q S
```

FIGURE 3B

pCMV Sequence and Restriction Site

```
Hind III, EcoR V, EcoR I, Pst I, Ava I, Nci I, Nci I, Sma I, BamH I, Spe I, Xba I, BsrB I, Not I, Hae III, Sac II, BstX I, Sac I

AAGCTTGATATCGAATTCCTGCAGCCCGGGGGATCCACTAGTTCTAGAGCGGCCGCCACCGCGGTGGAGCTCCAGCTTTT  80
TTCGAACTATAGCTTAAGGACGTCGGGCCCCCTAGGTGATCAAGATCTCGCCGGCGGTGGCGCCACCTCGAGGTCGAAAA

K L D I E F L Q P G G S T S S R A A A T A V E L Q L L
 S L I S N S C S P G D P L V L E R P P P R W S S S F
Q A . Y R I P A A R G I H . F . S G R H R G G A P A F

L S S I S N R C G P P D V L E L A A A V A T S S W S K
 L K I D F E Q L G P S G S T R S R G G G R H L E L K Q
 A Q Y R I G A A R P I W . N . L P R W R P P A G A K

BssH II

GTTCCCTTTAGTGAGGGTTAATTGCGCGCTAGAGGATCTTTGTGAAGGAACCTTACTTCTGTGGTGTGACATAATTGGAC  160
CAAGGGAAATCACTCCCAATTAACGCGCGATCTCCTAGAAACACTTCCTTGGAATGAAGACACCACACTGTATTAACCTG

 F P L V R V N C A L E D L C E G T L L L W C D I I G
C S L . . G L I A R . R I F V K E P Y F C G V T . L D
V P F S E G . L R A R G S L . R N L T S V V . H N W T

N G K T L T L Q A S S S R Q S P V K S R H H S M I P C
 E R . H P N I A R . L I K T F S G . K Q P T V Y N S
T G K L S P . N R A L P D K H L F R V E T T H C L Q V

Dra I

AAACTACCTACAGAGATTTAAAGCTCTAAGGTAAATATAAAATTTTTAAGTGTATAATGTGTTAAACTACTGATTCTAAT  240
TTTGATGGATGTCTCTAAATTTCGAGATTCCATTTATATTTTAAAAATTCACATATTACACAATTTGATGACTAAGATTA

Q T T Y R D L K L . G K Y K I F K C I M C . T T D S N
K L P T E I . S S K V N I K F L S V . C V K L L I L I
 N Y L Q R F K A L R . I . N F . V Y N V L N Y . F .

 V V . L S K F S . P L Y L I K L H I I H . V V S E L
L S G V S I . L E L T F I F N K L T Y H T L S S I R I
F . R C L N L A R L Y I Y F K . T Y L T N F . Q N . N

TGTTTGTGTATTTTAGATTCCAACCTATGGAACTGATGAATGGGAGCAGTGGTGGAATGCCTTTAATGAGGAAAACCTGT  320
ACAAACACATAAAATCTAAGGTTGGATACCTTGACTACTTACCCTCGTCACCACCTTACGGAAATTACTCCTTTTGGACA

C L C I L D S N L W N . . M G A V V E C L . . G K P V
 V C V F . I P T Y G T D E W E Q W W N A F N E E N L
L F V Y F R F Q P M E L M N G S S G G M P L M R K T C

Q K H I K S E L R H F Q H I P A T T S H R . H P F G T
 T Q T N . I G V . P V S S H S C H H F A K L S S F R N
 N T Y K L N W G I S S I F P L L P P I G K I L F V Q
```

FIGURE 3A

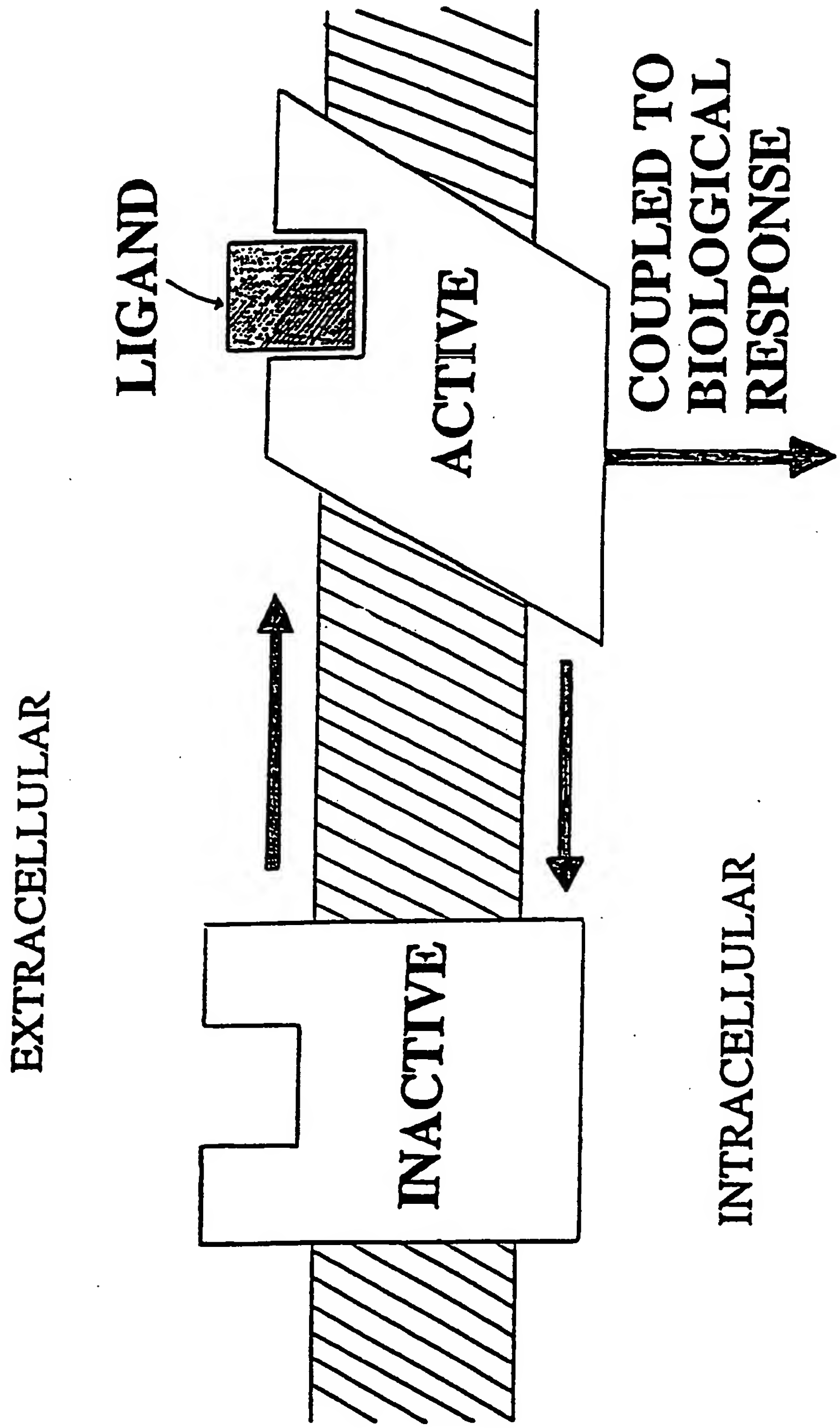
LIGAND
ACTIVE
COUPLED TO
BIOLOGICAL
RESPONSE
EXTRACELLULAR
INACTIVE
INTRACELLULAR

FIGURE 2

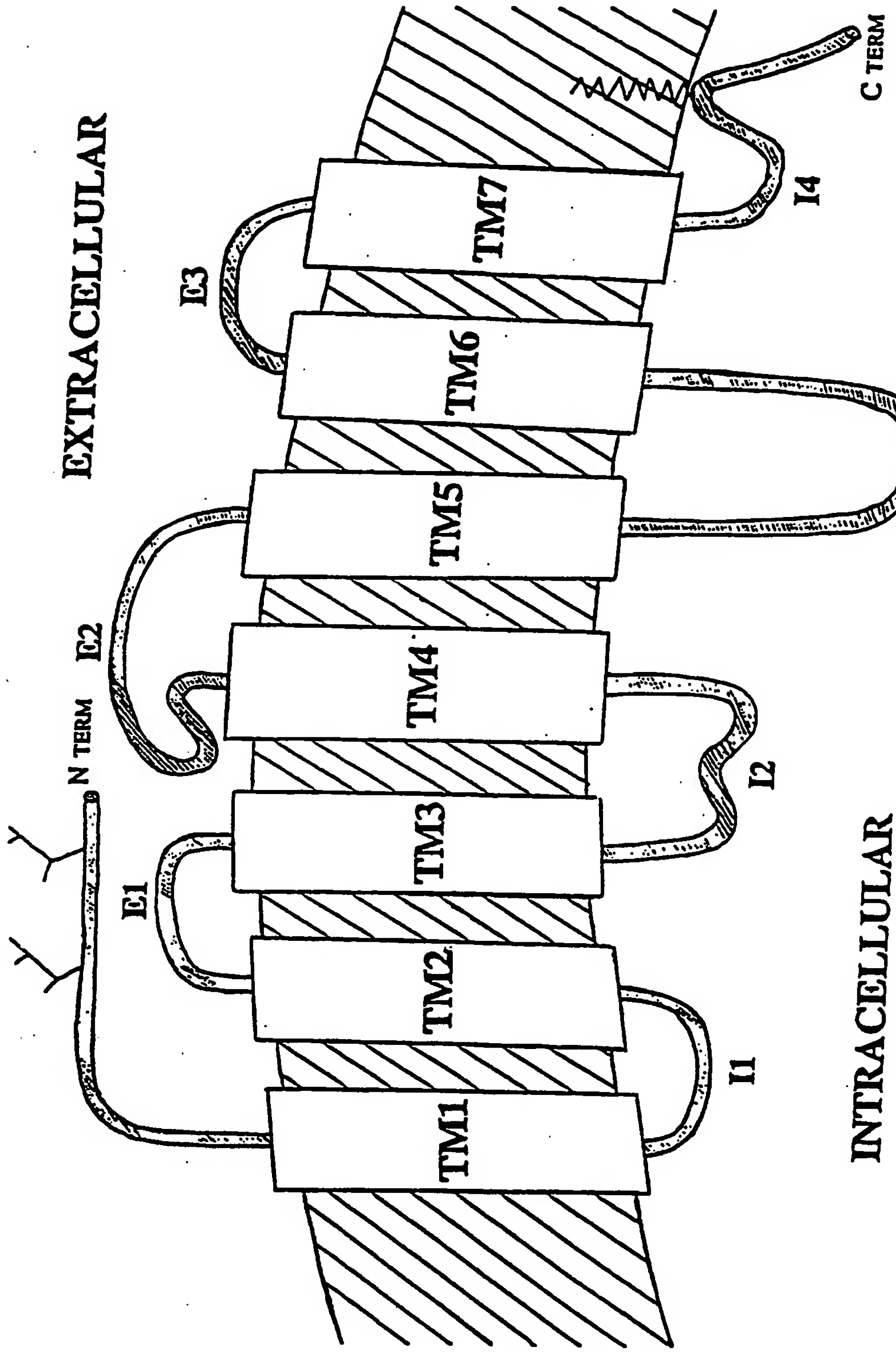
EXTRACELLULAR
N TERM
E1
E2
E3
TM1
TM2
TM3
TM4
TM5
TM6
TM7
C TERM
I1
I2
I3
I4
INTRACELLULAR

FIGURE 1

constitutively activated human G protein coupled receptor ("GPCR"), said GPCR comprising a transmembrane-6 region and an intracellular loop-3 region, comprising the steps of:

(a) selecting an endogenous human GPCR;

(b) identifying a proline residue within the transmembrane-6 region of the GPCR of step (a);

(c) identifying, in a carboxy-terminus to amino-terminus direction, the endogenous, 16^{th} amino acid residue from the proline residue of step (b);

(d) altering the endogenous amino acid of step (c) to a non-endogenous lysine residue;

(e) confirming that the non-endogenous GPCR of step (d) is constitutively active;

(f) contacting a candidate compound with the non-endogenous, constitutively-activated GPCR of step (e); and

(g) determining, by measurement of the compound efficacy at said contacted receptor, whether said compound is an inverse agonist of said receptor.

42. An inverse agonist directly identified by the method of claim 37.

43. A composition comprising an inverse agonist of claim 38.

(b) identifying a proline residue within the transmembrane-6 region of the GPCR of step (a);

(c) identifying, in a carboxy-terminus to amino-terminus direction, the endogenous,16^{th} amino acid residue from the proline residue of step (b);

(d) altering the endogenous amino acid of step (c) to a non-endogenous amino acid;

(e) confirming that the non-endogenous GPCR of step (d) is constitutively active;

(f) contacting a candidate compound with the non-endogenous, constitutively-activated GPCR of step (e); and

(g) determining, by measurement of the compound efficacy at said contacted receptor, whether said compound is an inverse agonist, agonist or partial agonist of said receptor.

33. The method of claim 32 wherein the non-endogenous amino acid of step (d) is lysine.

34. A compound directly identified by the method of claim 32.

35. The method of claim 32 wherein the directly identified compound is an inverse agonist.

36. The method of claim 32 wherein the directly identified compound is an agonist.--

37. The method of claim 32 wherein the directly identified compound is a partial agonist.

38. A composition comprising the inverse agonist of claim 35.

39. A composition comprising the agonist of claim 36.

40. A composition comprising the partial agonist of claim 37.

41. A method for directly identifying an inverse agonist to a non-endogenous,

algorithmic approach comprising the steps of:

(a) selecting an endogenous human GPCR comprising a proline residue in the transmembrane-6 region;

(b) identifying, by counting 16 amino acid residues from the proline residue of step (a), in a carboxy-terminus to amino-terminus direction, an endogenous amino acid residue;

(c) altering the identified amino acid residue of step (b) to a non-endogenous amino acid residue to create a non-endogenous version of the endogenous human GPCR; and

(d) determining if the non-endogenous version of the endogenous human GPCR of step (c) is constitutively active.

29. The algorithmic approach of claim 28 wherein the amino acid residue that is two residues from said proline residue in the transmembrane 6 region, in a carboxy-terminus to amino-terminus direction, is tryptophan.

30. A constitutively active, non-endogenous human GPCR produced by the algorithmic approach of claim 28.

31. A constitutively active, non-endogenous human GPCR produced by the algorithmic approach of claim 29.

32. A method for directly identifying a compound selected from the group consisting of inverse agonists, agonists and partial agonists to a non-endogenous, constitutively activated human G protein coupled receptor, said receptor comprising a transmembrane-6 region and an intracellular loop-3 region, comprising the steps of:

(a) selecting an endogenous human GPCR;

comprising a transmembrane 6 region and an intracellular loop 3 region, which endogenous amino acid, when altered to a non-endogenous amino acid, constitutively activates said human GPCR, comprising the following steps:

(a) identifying an endogenous proline residue within the transmembrane 6 region of a human GPCR;

(b) identifying, by moving in a direction of the carboxy-terminus region of said GPCR towards the amino-terminus region of said GPCR, the endogenous, 16th amino acid residue from said proline residue;

(c) altering the endogenous residue of step (b) to a non-endogenous amino acid residue to create a non-endogenous version of an endogenous human GPCR; and

(d) determining whether the non-endogenous human GPCR of step (c) is constitutively active.

25. The method of claim 24 wherein the amino acid residue that is two residues from said proline residue in the transmembrane 6 region, in a carboxy-terminus to amino-terminus direction, is tryptophan.

26. A constitutively active, non-endogenous human GPCR produced by the process of claim 24.

27. A constitutively active, non-endogenous human GPCR produced by the process of claim 25.

28. An algorithmic approach for creating a non-endogenous, constitutively active version of an endogenous human G protein coupled receptor (GPCR), said endogenous GPCR comprising a transmembrane 6 region and an intracellular loop 3 region, the

amino acid residue other than a proline residue.

15. The nucleic acid sequence of claim 12 wherein X_{codon} encodes a non-endogenous amino acid selected from the group consisting of lysine, histidine, arginine and alanine, excepting that when the endogenous amino acid in position X of said endogenous human GPCR is lysine, X_{codon} encodes an amino acid selected from the group consisiting of histidine, arginine and alanine.

16. The nucleic acid sequence of claim 13 wherein X_{codon} encodes a non-endogenous lysine amino acid excepting that when the endogenous amino acid in position X of said endogenous human GPCR is lysine, X_{codon} encodes an amino acid selected from the group consisiting of histidine, arginine and alanine.

17. The nucleic acid sequence of claim 12 wherein X_{codon} is selected from the group consisting of AAA, AAG, GCA, GCG, GCC and GCU.

18. The nucleic acid sequence of claim 12 wherein X_{codon} is selected from the group consisting of AAA and AAG.

19. The nucleic acid sequence of claim 12 wherein P^{codon} is selected from the group consisting of CCA, CCC, CCG and CCU, and X_{codon} is selected from the group consisting of AAA and AAG.

20. A vector comprising the nucleic acid sequence of claim 12.

21. A plasmid comprising the nucleic acid sequence of claim 12.

22. A host cell comprising the nucleic acid sequence of claim 21.

23. The nucleic acid sequence of claim 12 in a purified and isolated form.

24. A method for selecting for alteration an endogenous amino acid residue within the third intracellular loop of a human G protein-coupled receptor ("GPCR"), said receptor

nucleic acid sequence region transversing the transmembrane-6 (TM6) and intracellular loop-3 (IC3) regions of the orphan GPCR:

$$3'\text{-}P^{codon}\,(\text{AA-codon})_{15}\,X_{codon}\text{-}5'$$

wherein:

(1) P^{codon} is a nucleic acid encoding region within the TM6 region of the non-endogenous GPCR, where P^{codon} encodes an amino acid selected from the group consisting of (i) the endogenous GPCR proline residue, and (ii) a non-endogenous amino acid residue other than proline;

(2) $(\text{AA-codon})_{15}$ are 15 codons encoding 15 amino acid residues selected from the group consisting of (a) the 15 endogenous amino acid residues of the endogenous orphan GPCR, (b) 15 non-endogenous amino acid residues, and (c) a combination of 15 amino acid residues, the combination comprising at least one endogenous amino acid residue of the endogenous orphan GPCR and at least one non-endogenous amino acid residue, excepting that none of the 15 endogenous amino acid residues that are positioned within the TM6 region of the orphan GPCR is proline; and

(3) X_{codon} is a nucleic acid encoding region residue located within the IC3 region of said non-endogenous human GPCR, where X_{codon} encodes a non-endogenous amino acid.

13. The nucleic acid sequence of claim 12 wherein P^{codon} encodes an endogenous proline residue.

14. The nucleic acid sequence of claim 12 wherein P^{codon} encodes a non-endogenous

residue.

3. The non-endogenous human GPCR of claim1 wherein P^1 is a non-endogenous amino acid residue other than a proline residue.

4. The non-endogenous human GPCR of claim 1 wherein AA_{15} are the 15 endogenous amino acid residues of the endogenous GPCR.

5. The non-endogenous human GPCR of claim 1 wherein X is selected from the group consisting of lysine, hisitidine, arganine and alanine residues, excepting that when the endogenous amino acid in position X of said endogenous human GPCR is lysine, X is selected from the group consisting of histidine, arginine and alanine.

6. The non-endogenous human GPCR of claim 1 wherein X is a lysine residue, excepting that when the endogenous amino acid in position X of said endogenous human GPCR is lysine, X is an amino acid other than lysine.

7. The non-endogenous human GPCR of claim 4 wherein X is a lysine residue, excepting that when the endogenous amino acid in position X of said endogenous human GPCR is lysine, X is an amino acid other than lysine.

8. The non-endogenous, human GPCR of claim 1 wherein P^1 is a proline residue and X is a lysine residue, excepting that when the endogenous amino acid in position X of said endogenous human GPCR is lysine, X is an amino acid other than lysine.

9. A host cell comprising the non-endogenous human GPCR of claim 1.

10. The material of claim 9 wherein said host cell is of mammalian origin.

11. The non-endogenous human GPCR of claim 1 in a purified and isolated form.

12. A nucleic acid sequence encoding a constitutively active, non-endogenous version of an endogenous human orphan G protein-coupled receptor (GPCR) comprising the following

CLAIMS

What is claimed is:

1. A constitutively active, non-endogenous version of an endogenous human orphan G protein-coupled receptor (GPCR) comprising the following amino acid residues (carboxy-terminus to amino-terminus orientation) transversing the transmembrane-6 (TM6) and intracellular loop-3 (IC3) regions of the non-endogenous GPCR:

$$P^1\ AA_{15}\ X$$

wherein:

(1) P^1 is an amino acid residue located within the TM6 region of the non-endogenous GPCR, where P^1 is selected from the group consisting of (i) the endogenous orphan GPCR proline residue, and (ii) a non-endogenous amino acid residue other than proline;

(2) AA_{15} are 15 amino acid residues selected from the group consisting of (a) the 15 endogenous amino acid residues of the endogenous orphan GPCR, (b) 15 non-endogenous amino acid residues, and (c) a combination of 15 amino acid residues, the combination comprising at least one endogenous amino acid residue of the endogenous orphan GPCR and at least one non-endogenous amino acid residue, excepting that none of the 15 endogenous amino acid residues that are positioned within the TM6 region of the GPCR is proline; and

(2) X is a non-endogenous amino acid residue located within the IC3 region of said non-endogenous GPCR.

2. The non-endogenous human GPCR of claim 1 wherein P^1 is the endogenous proline

candidate compounds of potential therapeutic relevance, it is noted that inverse agonists are useful in the treatment of diseases and disorders where a particular human GPCR is over-expressed, whereas agonists or partial agonists are useful in the treatment of diseases and disorders where a particular human GPCR is under-expressed.

As desired, more detailed, cellular localization of the recepotrs, using techniques well-known to those in the art (*e.g.*, in-situ hybridization) can be utilized to identify particualr cells within these tissues where the receptor of interest is expressed.

It is intended that each of the patents, applications, and printed publications mentioned in this patent document be hereby incorporated by reference in their entirety.

As those skilled in the art will appreciate, numerous changes and modifications may be made to the preferred embodiments of the invention without departing from the spirit of the invention. It is intended that all such variations fall within the scope of the invention.

Although a variety of expression vectors are available to those in the art, for purposes of utilization for both the endogenous and non-endogenous human GPCRs, it is most preferred that the vector utilized be pCMV. This vector has been deposited with the American Type Culture Collection (ATCC) on October 13, 1998 (10801 University Blvd., Manassas, VA 20110-2209 USA) under the provsions of the Budapest Treaty for the International Recognition of the Deposit of Microorganisms for the Purpose of patent Procedure. The vector was tested by the ATCC on _________, 1998 and determined to be viable on ___________, 1998. The ATCC has assigned the following deposit number to pCMV: _________.

GPR4	Broad; highest in Heart, Lung, Adrenal, Thyroid, Spinal Cord
GPR5	Placenta, Thymus, Fetal Thymus Lesser levels in spleen, fetal spleen
GPR7	Liver, Spleen, Spinal Cord, Placenta
GPR8	No expression detected
GPR9-6	Thymus, Fetal Thymus Lesser levels in Small Intestine
GPR18	Spleen, Lymph Node, Fetal Spleen, Testis
GPR20	Broad
GPR21	Broad; very low abundance
GPR22	Heart, Fetal Heart Lesser levels in Brain
GPR30	Stomach
GPR31	Broad
BLR1	Spleen
CEPR	Stomach, Liver, Thyroid, Putamen
EBI1	Pancreas Lesser levels in Lymphoid Tissues
EBI2	Lymphoid Tissues, Aorta, Lung, Spinal Cord
ETBR-LP2	Broad; Brain Tissue
GPCR-CNS	Brain Lesser levels in Testis, Placenta
GPR-NGA	Pituitary Lesser levels in Brain
H9	Pituitary
HB954	Aorta, Cerebellum Lesser levels in most other tissues
HM74	Spleen, Leukocytes, Bone marrow, Mammary Glands, Lung, Trachea
MIG	Low levels in Kidney, Liver, Pancreas, Lung, Spleen
ORG1	Pituitary, Stomach, Placenta
V28	Brain, Spleen, Peripheral Leukocytes

Based upon the foregoing information, it is noted that human GPCRs can also be assessed for distribution in diseased tissue; comparative assessments between "normal" and diseased tissue can then be utilized to determine the potential for over-expression or under-expression of a particular receptor in a diseased state. In those circumstances where it is desirable to utilize the non-endogenous versions of the human GPCRs for the purpose of screening to directly identify

MIG (I230K)	39 (cAMP)
Serotonin $5HT_{2A}$ (C322K)	33.2 (IP_3)
Serotonin $5HT_{2C}$ (S310K)	39.1(IP_3)

Example 6

Tissue Distribution of Endogenous Orphan GPCRs

Using a commercially available human-tissue dot-blot format, endogenous orphan GPCRs were probed for a determination of the areas where such receptors are localized. Except as indicate below, the entire receptor cDNA (radiolabelled) was used as the probe: radiolabeled probe was generated using the complete receptor cDNA (excised from the vector) using a Prime-It II™ Random Primer Labeling Kit (Stratagene, #300385), according to manufacturer's instructions. A human RNA Master Blot™ (Clontech, #7770-1) was hybridized with the GPCR radiolabeled probe and washed under stringent conditions according manufacturer's instructions. The blot was exposed to Kodak BioMax Autoradiography film overnight at -80°C.

Representative dot-blot format results are presented in Figure 8 for GPR1 (8A), GPR30 (8B), and APJ (8C), with results being summarized for all receptors in Table F

Table F

GPCR	Tissue Distribution (highest levels, relative to other tissues in the dot-blot)
GPR1	Placenta, Ovary, Adrenal

can be further examined, we believe that this difference is itself useful in differentiating between the endogenous and non-endogenous versions of the human $5HT_{2A}$ receptor.

D. Result Summary

The results for the GPCRs tested are set forth in Table E where the Per-Cent Increase indicates the percentage difference in results observed for the non-endogenous GPCR as compared to the endogenous GPCR; these values are followed by parenthetical indications as to the type of assay utilized. Additionally, the assay sytem utilized is parenthetically listed (and, in cases where different Host Cells were used, both are listed). As these results indicate, a variety of assays can be utilized to determine constitutive activity of the non-endogenous versions of the human GPCRs. Those skilled in the art, based upon the foregoing and with reference to information available to the art, are creditied with theability to selelect and/ot maximize a particular assay approach that suites the particualr needs of theinvestigator.

Table E

Receptor Identifier (Codon Mutation)	Per-Cent Difference
GPR17 (V234K)	74.5 (CRE-Luc)
GPR30 (L258K)	71.6 (CREB)
APJ (L247K)	49.0 (GTPγS)
ETBR-LP2 (N358K)	48.4(AP1-Luc - 293) 61.1(AP1-Luc – 293T)
GHSR (V262K)	58.9(CREB – 293) 35.6(CREB – 293T)

and the cells were incubated for 16-18 hrs o/n at 37°C/5%CO_2. On Day 4 the cells were washed with 0.5 ml PBS and 0.45 ml of assay medium was added containing inositol-free/serum free media 10 μM pargyline 10 mM lithium chloride or 0.4 ml of assay medium and 50 ul of 10x ketanserin (ket) to final concentration of 10μM. The cells were then incubated for 30 min at 37°C. The cells were then washed with 0.5 ml PBSand 200 ul of fresh/icecold stop solution (1M KOH; 18 mM Na-borate; 3.8 mM EDTA) was added/well. The solution was kept on ice for 5-10 min or until cells were lysed and then neutralized by 200 μl of fresh/ice cold neutralization sol. (7.5 % HCL). The lysate was then transferred into 1.5 ml eppendorf tubes and 1 ml of chloroform/methanol (1:2) was added/tube. The solution was vortexed for 15 sec and the upper phase was applied to a Biorad AG1-X8 anion exchange resin (100-200 mesh). Firstly, the resin was washed with water at 1:1.25 W/V and 0.9 ml of upper phase was loaded onto the column. The column was washed with 10 mls of 5 mM myo-inositol and 10 ml of 5 mM Na-borate/60mM Na-formate. The inositol tris phosphates were eluted into scintillation vials containing 10 ml of scintillation cocktail with 2 ml of 0.1 M formic acid/ 1 M ammonium formate. The columns were regenerated by washing with 10 ml of 0.1 M formic acid/3M ammonium formate and rinsed twice with dd H_2O and stored at 4°C in water.

Figure 7 provides an illustration of IP3 production from the human 5-HT_{2A} receptor that incorporates the C322K mutation. While these results evidence that the Proline Mutation Algorithm approach constitutively activates this receptor, for purposes of using such a receptor for screening for identification of potential therapeutics, a more robust difference would be preferred. However, because the activated receptor can be utilized for understanding and elucidating the role of constitutive activation and for the identification of compounds that

transfection efficiency between samples) were combined in a calcium phosphate precipitate as per the Kit's instructions. Half of the precipitate was equally distributed over 3 wells in a 96-well plate, kept on the cells overnight, and replaced with fresh medium the following morning. Forty-eight (48) hr after the start of the transfection, cells were treated and assayed for luciferase activity as set forth with resepct to the GPR30 system, above. This assay was used with respect to GHSR.

2. AP1 reporter assay (Gq-associated receptors)

Ae method to detect Gq stimulation depends on the known property of Gq-dependent phospholipase C to cause the activation of genes containing AP1 elements in their promoter. A Pathdetect AP-1 cis-Reporting System (Stratagene, Catalogue # 219073) was utilized following the protocl set forth above with respect to the CREB reporter assay, except that the components of the calcium phosphate precipitate were 410 ng pAP1-Luc, 80 ng receptor expression plasmid, and 20 ng CMV-SEAP. This assay was used with respect to ETBR-LP2

C. Intracellular IP3 Accumulation Assay

On day 1, cells comprising the serotonin receptors (endogenous and mutated) were plated onto 24 well plates, usually $1x10^5$ cells/well. On day 2 cells were transfected by firstly mixing 0.25ug DNA in 50 ul serumfree DMEM/well and 2 ul lipofectamine in 50 µl serumfree DMEM/well. The solutions were gently mixed and incubated for 15-30 min at room temperature. Cells were washed with 0.5 ml PBS and 400 µl of serum free media was mixed with the transfection media and added to the cells. The cells were then incubated for 3-4 hrs at 37°C/5%CO_2 and then the transfection media was removed and replaced with 1ml/well of regular growth media. On day 3 the cells were labeled with ^{3}H-myo-inositol. Briefly, the media was removed the cells were washed with 0.5 ml PBS. Then 0.5 ml inositol-free/serumfree media (GIBCO BRL) was added/well with 0.25 µCi of ^{3}H-myo-inositol / well

11 ml Detection Buffer) were prepared and maintained in accordance with the manufacturer's instructions. Assay Buffer was prepared fresh for screening and contained 20mM HEPES, pH 7.4, 10mM $MgCl_2$, 20mM (Sigma), 0.1 units/ml creatine phosphokinase (Sigma), 50 μM GTP (Sigma), and 0.2 mM ATP (Sigma); Assay Buffer can be stored on ice until utilized. The assay was initiated by addition of 50ul of assay buffer followed by addition of 50ul of membrane suspension to the NEN Flash Plate. The resultant assay mixture is incubated for 60 minutes at room temperature followed by addition of 100ul of detection buffer. Plates are then incubated an additional 2-4 hours followed by counting in a Wallac MicroBeta scintillation counter. Values of cAMP/well are extrapolated from a standard cAMP curve which is contained within each assay plate. The foregoing assay was utilized with respect to analysis of MIG.

B. Reporter-Based Assays

1. CREB Reporter Assay (Gs-associated receptors)

A method to detect Gs stimulation depends on the known property of the transcription factor CREB, which is activated in a cAMP-dependent manner. A PathDetect CREB trans-Reporting System (Stratagene, Catalogue # 219010) was utilized to assay for Gs coupled activity in 293 or 293T cells. Cells were transfected with the plasmids components of this above system and the indicated expression plasmid encoding endogenous or mutant receptor using a Mammalian Transfection Kit (Stratagene, Catalogue #200285) according to the manufacurer's instructions. Briefly, 400 ng pFR-Luc (luciferase reporter plasmid containing Gal4 recognition sequences), 40 ng pFA2-CREB (Gal4-CREB fusion protein containing the Gal4 DNA-binding domain), 80 ng CMV-receptor expression plasmid (comprising the receptor) and 20 ng CMV-SEAP (secreted alkaline phosphatase expression plasmid; alkaline phosphatase activity is measured in the media of transfected cells to control for variations in

ligand is centrifuged to the bottom of the well, the scintistrip label comes into proximity with the radiolabeled ligand resulting in activation and detection.

Representative results of graph comparing Control (pCMV), Endogenous APJ and Non-Endogenous APJ, based upon the foregoing protocol, are set forth in Figure 6.

2. Adenylyl Cyclase

A Flash Plate™ Adenylyl Cyclase kit (New England Nuclear; Cat. No. SMP004A) designed for cell-based assays was modified for use with crude plasma membranes. The Flash Plate wells contain a scintillant coating which also contains a specific antibody recognizing cAMP. The cAMP generated in the wells was quantitated by a direct competition for binding of radioactive cAMP tracer to the cAMP antibody. The following serves as a brief protocol for the measurement of changes in cAMP levels in membranes that express the receptors.

Transfected cells were harvested approximately three days after transfection. Membranes were prepared by homogenization of suspended cells in buffer containing 20mM HEPES, pH 7.4 and 10mM $MgCl_2$. Homogenization was performed on ice using a Brinkman Polytron™ for approximately 10 seconds. The resulting homogenate was centrifuged at 49,000 X g for 15 minutes at 4°C. The resulting pellet was then resuspended in buffer containing 20mM HEPES, pH 7.4 and 0.1 mM EDTA, homogenized for 10 seconds, followed by centrifugation at 49,000 X g for 15 minutes at 4°C. The resulting pellet can be stored at -80°C until utilized. On the day of measurement, the membrane pellet was slowly thawed at room temperature, resuspended in buffer containing 20mM HEPES, pH 7.4 and 10mM $MgCL_2$ (these amounts can be optimized, although the values listed herein are prefereed), to yield a final protein concentration of 0.60mg/ml (the resuspended membranes were placed on ice until use).

cAMP standards and Detection Buffer (comprising 2 µCi of tracer [^{125}I cAMP (100 µl] to

to drug discovery at all G protein-coupled receptors.

The [^{35}S]GTPγS assay was incubated in 20 mM HEPES and between 1 and about 20mM $MgCl_2$ (this amount can be adjusted for optimization of results, although 20mM is preferred) pH 7.4, binding buffer with between about 0.3 and about 1.2 nM [^{35}S]GTPγS (this amount can be adjusted for optimization of results, although 1.2 is preferred) and 12.5 to 75 μg membrane protein (*e.g.* COS-7 cells expressing the receptor; this amount can be adjusted for optimization, although 75μg is preferred) and 1 μM GDP (this amount can be changed for optimization) for 1 hour. Wheatgerm agglutinin beads (25 μl; Amersham) were then added and the mixture was incubated for another 30 minutes at room temperature. The tubes were then centrifuged at 1500 x g for 5 minutes at room temperature and then counted in a scintillation counter.

A less costly but equally applicable alternative has been identified which also meets the needs of large scale screening. Flash plates™ and Wallac™ scintistrips may be utilized to format a high throughput [^{35}S]GTPγS binding assay. Furthermore, using this technique, the assay can be utilized for known GPCRs to simultaneously monitor tritiated ligand binding to the receptor at the same time as monitoring the efficacy via [^{35}S]GTPγS binding. This is possible because the Wallac beta counter can switch energy windows to look at both tritium and ^{35}S-labeled probes. This assay may also be used to detect other types of membrane activation events resulting in receptor activation. For example, the assay may be used to monitor ^{32}P phosphorylation of a variety of receptors (both G protein coupled and tyrosine kinase receptors). When the membranes are centrifuged to the bottom of the well, the bound [^{35}S]GTPγS or the ^{32}P-phosphorylated receptor will activate the scintillant which is coated of the wells. Scinti® strips (Wallac) have been used to demonstrate this principle. In addition, the assay also has utility for measuring ligand binding to receptors using radioactively labeled ligands. In a similar manner, when the radiolabeled bound

expression to about 104.1 relative light units. Co-transfection of endogenous GPR6 with non-endogenous GPR30 (L258K), at the same ratio, drastically decreased the expression, which is evident at about 18.2 and 29.5 relative light units, respectively. Similar results were observed with respect to GPR17 with respect to co-transfection with GPR3, as set forth in Figure 5.

Example 3
ASSAYS FOR DETERMINATION OF CONSTITUTIVE ACTIVITY OF NON-ENDOGENOUS GPCRS

A. Membrane Binding Assays

1. [^{35}S]GTPγS Assay

When a G protein-coupled receptor is in its active state, either as a result of ligand binding or constitutive activation, the receptor couples to a G protein and stimulates the release of GDP and subsequent binding of GTP to the G protein. The alpha subunit of the G protein-receptor complex acts as a GTPase and slowly hydrolyzes the GTP to GDP, at which point the receptor normally is deactivated. Constitutively activated receptors continue to exchange GDP for GTP. The non-hydrolyzable GTP analog, [^{35}S]GTPγS, can be utilized to demonstrate enhanced binding of [^{35}S]GTPγS to membranes expressing constitutively activated receptors. The advantage of using [^{35}S]GTPγS binding to measure constitutive activation is that: (a) it is generically applicable to all G protein-coupled receptors; (b) it is proximal at the membrane surface making it less likely to pick-up molecules which affect the intracellular cascade.

The assay utilizes the ability of G protein coupled receptors to stimulate [^{35}S]GTPγS binding to membranes expressing the relevant receptors. The assay can, therefore, be used in the direct identification method to screen candidate compounds to known, orphan and constitutively activated G protein-coupled receptors. The assay is generic and has application

reporter plasmid (*see* below), 50ng of pCMV comprising endogenous receptor or non-endogenous receptor or pCMV alone, and 10ng of a GPRS expression plasmid (GPRS in pcDNA3 (Invitrogen)). The 8XCRE-Luc reporter plasmid was prepared as follows: vector SRIF-β-gal was obtained by cloning the rat somatostatin promoter (-71/+51) at BglV-HindIII site in the pβgal-Basic Vector (Clontech). Eight (8) copies of cAMP response element were obtained by PCR from an adenovirus template AdpCF126CCRE8 (*see* 7 Human Gene Therapy 1883 (1996)) and cloned into the SRIF-β-gal vector at the Kpn-BglV site, resulting in the 8xCRE-β-gal reporter vector. The 8xCRE-Luc reporter plasmid was generated by replacing the beta-galactosidase gene in the 8xCRE-β-gal reporter vector with the luciferase gene obtained from the pGL3-basic vector (Promega) at the HindIII-BamHI site. Following 30min. incubation at room temperature, the DNA/lipid mixture was diluted with 400 µl of DMEM and 100µl of the diluted mixture was added to each well. 100 µl of DMEM with 10% FCS were added to each well after a 4hr incubation in a cell culture incubator. The next morning the transfected cells were changed with 200 µl/well of DMEM with 10% FCS. Eight (8) hours later, the wells were changed to 100 µl /well of DMEM without phenol red, after one wash with PBS. Luciferase activity were measured the next day using the LucLite™ reporter gene assay kit (Packard) following manufacturer instructions and read on a 1450 MicroBeta™ scintillation and luminescence counter (Wallac).

Figure 4 evidences that constitutively active GPR30 inhibits GPR6-mediated activation of CRE-Luc reporter in 293 cells. Luciferase was measured at about 4.1 relative light units in the expression vector pCMV. Endogenous GPR30 expressed luciferase at about 8.5 relative light units, whereas the non-endogenous, constitutively active GPR30 (L258K), expressed luciferase at about 3.8 and 3.1 relative light units, respectively. Co-transfection of endogenous GPR6 with endogenous GPR30, at a 1:4 ratio, drastically increased luciferase

skilled artisan to determine an appropriate reporter plasmid for a particular gene expression based primarily upon the particular need of the artisan. Although a variety of cells are available for expression, mammalian cells are most preferred, and of these types, 293 cells are most preferred. 293 cells were transfected with the reporter plasmid pCRE-Luc/GPR6 and non-endogenous, constitutively activated GPR30 using a Mammalian Transfection™ Kit (Stratagene, #200285) $CaPO_4$ precipitation protocol according to the manufacturer's instructions (*see*, 28 Genomics 347 (1995) for the published endogenous GPR6 sequence). The precipitate contained 400ng reporter, 80ng CMV-expression plasmid (having a 1:4 GPR6 to endogenous GPR30 or non-endogenous GPR30 ratio) and 20ng CMV-SEAP (a transfection control plasmid encoding secreted alkaline phosphatase). 50% of the precipitate was split into 3 wells of a 96-well tissue culture dish (containing $4X10^4$ cells/well); the remaining 50% was discarded. The following morning, the media was changed. 48 hr after the start of the transfection, cells were lysed and examined for luciferase activity using a Luclite™ Kit (Packard, Cat. # 6016911) and Trilux 1450 Microbeta™ liquid scintillation and luminescence counter (Wallac) as per the vendor's instructions. The data were analyzed using GraphPad Prism 2.0a (GraphPad Software Inc.).

With respect to GPR17, which has also been determined to be Gi-linked, a modification of the foregoing approach was utilized, based upon, *inter alia*, use of another Gs-linked endogenous receptor, GPR3 (*see* 23 Genomics 609 (1994) and 24 Genomics 391 (1994)). Most preferably, 293 cells are utilized. These cells were plated-out on 96 well plates at a density of 2×10^4 cells per well and were transfected using Lipofectamine Reagent (BRL) the following day according to manufacturer instructions. A DNA/lipid mixture was prepared for each 6-well transfection as follows: 260ng of plasmid DNA in 100µl of DMEM were gently mixed with 2µl of lipid in 100µl of DMEM (the 260ng of plasmid DNA consisted of 200ng of a 8xCRE-Luc

Assays that detect cAMP can be utilized to determine if a candidate compound is *e.g.*, an inverse agonist to a Gs-associated receptor (*i.e.*, such a compound would decrease the levels of cAMP) or a Gi-associated receptor (or a Go-associated receptor) (*i.e.*, such a candidate compound would increase the levels of cAMP). A variety of approaches known in the art for measuring cAMP can be utilized; a preferred approach relies upon the use of anti-cAMP antibodies. Another approach, and most preferred, utilizes a whole cell second messenger reporter system assay. Promoters on genes drive the expression of the proteins that a particular gene encodes. Cyclic AMP drives gene expression by promoting the binding of a cAMP-responsive DNA binding protein or transcription factor (CREB) which then binds to the promoter at specific sites called cAMP response elements and drives the expression of the gene. Reporter systems can be constructed which have a promoter containing multiple cAMP response elements before the reporter gene, *e.g.*, β-galactosidase or luciferase. Thus, an activated receptor such as GPR6 causes the accumulation of cAMP which then activates the gene and expression of the reporter protein. Most preferably, 293 cells are co-transfected with GPR6 (or another Gs-linked receptor) and GPR30 (or another Gi-linked receptor) plasmids, preferably in a 1:1 ratio, most preferably in a 1:4 ratio. Because GPR6 is an endogenous, constitutively active receptor that stimulates the production of cAMP, GPR6 strongly activates the reporter gene and its expression. The reporter protein such as β-galactosidase or luciferase can then be detected using standard biochemical assays (Chen et al. 1995). Co-transfection of endogenous, constitutively active GPR6 with endogenous, non-constitutively active GPR30 evidences an increase in the luciferase reporter protein. Conversely, co-transfection of endogenous, constitutively active GPR6 with non-endogenous, constitutively active GPR30 evidences a drastic decrease in expression of luciferase. Several reporter plasmids are known and available in the art for measuring a second messenger assay. It is considered well within the

Irvine, CA); tube B was prepared by mixing 120μl lipofectamine (Gibco BRL) in 1.2ml serum free DMEM. Tubes A and B were then admixed by inversions (several times), followed by incubation at room temperature for 30-45min. The admixture is referred to as the "transfection mixture". Plated cells were washed with 1XPBS, followed by addition of 10ml serum free DMEM. 2.4ml of the transfection mixture was then added to the cells, followed by incubation for 4hrs at 37°C/5% CO_2. The transfection mixture was then removed by aspiration, followed by the addition of 25ml of DMEM/10% Fetal Bovine Serum. Cells were then incubated at 37°C/5% CO_2. After 72hr incubation, cells were then harvested and utilized for analysis.

1. Gi-Coupled Receptors: Co-Transfection with Gs-Coupled Receptors

In the case of GPR30, it has been determined that this receptor couples the G protein Gi. Gi is known to inhibit the enzyme adenylyl cyclase, which is necessary for catalyzing the conversion of ATP to cAMP. Thus, a non-endogenous, constitutively activated form of GPR30 would be expected to be associated with decreased levels of cAMP. Assay confirmation of a non-endogenous, constitutively activated form of GPR30 directly via measurement of decreasing levels of cAMP, while viable, can be preferably measured by cooperative use of a Gs-coupled receptor. For example, a receptor that is Gs-coupled will stimulate adenylyl cyclase, and thus will be associated with an increase in cAMP. The assignee of the present application has discovered that the orphan receptor GPR6 is an endogenous, constitutively activated GPCR. GPR6 couples to the Gs protein. Thus when co-transfected, one can readily verify that a putative GPR30-mutation leads to constitutive activation thereof: *i.e.*, an endogenous, constitutively activated GPR6/endogenous, non-constitutively activated GPR30 cell will evidence an elevated level of cAMP when compared with an endogenous, constitutively active GPR6/non-endogenous, constitutively activated GPR30 (the latter evidencing a comparatively lower level of cAMP).

introduces into the protocol a non-mammalian cell which may not (indeed, in the case of yeast, does not) include the receptor-coupling, genetic-mechanism and secretary pathways that have evolved for mammalian systems – thus, results obtained in non-mammalian cells, while of potential use, are not as preferred as that obtained from mammalian cells. Of the mammalian cells, COS-7, 293 and 293T cells are particularly preferred, although the specific mammalian cell utilized can be predicated upon the particular needs of the artisan.

Unless otherwise noted herein, the following protocol was utilized for the expression of the endogenous and non-endogenous human GPCRs. Table D lists the mammalian cell and number utilized (per 150mm plate) for GPCR expression.

Table D

Receptor Name (Endogenous or Non-Endogenous)	**Mammalian Cell** (Number Utilized)
GPR17	293 (2 x 10^4)
GPR30	293 (4 x 10^4)
APJ	COS-7 (5X10^6)
ETBR-LP2	293 (1 x 10^7) 293T (1 x 10^7)
GHSR	293 (1 x 10^7) 293T (1 x 10^7)
MIG	293 (1 x 10^7)
Serotonin $5HT_{2A}$	293T (1 x 10^7)
Serotonin $5HT_{2c}$	293T (1 x 10^7)

On day one, mammalian cells were plated out. On day two, two reaction tubes were prepared (the proportions to follow for each tube are per plate): tube A was prepared by mixing 20μg DNA (*e.g.*, pCMV vector; pCMV vector with endogenous receptor cDNA, and pCMV vector with non-endogenous receptor cDNA.) in 1.2ml serum free DMEM (Irvine Scientific,

the desired mutation. The sense strand sequence utilized had the following sequence:

5'-CTAGGGGCACCATGCAGGCTATCAACAATGAAAGAAAAGCTAAGAAAGTC-3' (SEQ. ID.NO.: 237)

and the antisense strand sequence utilized had the following sequence:

5'-CAAGGACTTTCTTAGCTTTTCTTTCATTGTTGATAGCCTGCATGGTGCCC-3' (SEQ. ID. NO.: 238)

d. GPR30

Prior to generating non-endogenous GPR30, several independent pCR2.1/GPR30 isolates were sequenced in their entirety in order to identify clones with no PCR-generated mutations. A clone having no mutations was digested with EcoR1 and the endogenous GPR30 cDNA fragment was transferred into the CMV-driven expression plasmid pCI-neo (Promega), by digesting pCI-Neo with EcoRI and subcloning the EcoRI-liberated GPR30 fragment from pCR2.1/GPR30, to generate pCI/GPR30. Thereafter, the leucine at codon 258 was mutated to a lysine using a Quick-Change™ Site-Directed Mutagenesis Kit (Stratagene, #200518), according to manufacturer's instructions, and the following primers:

5'-CGGCGGCAGAAGGCGAAACGCATGATCCTCGCGGT-3' (SEQ.ID.NO.: 239) and

5'-ACCGCGAGGATCATGCGTTTCGCCTTCTGC CGCCG-3' (SEQ.ID.NO.: 240)

Example 3

Receptor (Endogenous and Mutated) Expression

Although a variety of cells are available to the art for the expression of proteins, it is most preferred that mammalian cells be utilized. The primary reason for this is predicated upon practicalities, *i.e.*, utilization of, *e.g.*, yeast cells for the expression of a GPCR, while possible,

2. Alternate Mutation Approaches for Employment of the Proline Marker Algorithm: APJ; Serotonin $5HT_{2A}$; Serotonin $5HT_{2C}$; and GPR30

Although the above site-directed mutagenesis approach is particularly preferred, other approaches can be utilized to create such mutations; those skilled in the art are readily credited with selecting approaches to mutating a GPCR that fits within the particular needs of the artisan.

a. APJ

Preparation of the non-endogenous, human APJ receptor was accomplished by mutating L247K. Two oligonucleotides containing this mutation were synthesized:

5'- GGCTTAAGAGCATCATCGTGGTGCTGGTG-3' (SEQ.ID.NO.: 233)

5'-GTCACCACCAGCACCACGATGATGCTCTTAAGCC-3' (SEQ.ID.NO.: 234)

The two oligonucleotides were annealed and used to replace the NaeI-BstEII fragment of human, endogenous APJ to generate the non-endogenous, version of human APJ.

b. Serotonin $5HT_{2A}$

cDNA containing the point mutation C322K was constructed by utilizing the restriction enzyme site Sph I which encompasses amino acid 322. A primer containing the C322K mutation:

5'-CAAAGAAAGTACTGGGCATCGTCTTCTTCCT-3' (SEQ.ID.NO: 235)

was used along with the primer from the 3' untranslated region of the receptor:

5'-TGCTCTAGATTCCAGATAGGTGAAAA CTTG-3' (SEQ.ID.NO.: 236)

to perform PCR (under the conditions described above). The resulting PCR fragment was then used to replace the 3' end of endogenous $5HT_{2A}$ cDNA through the T4 polymerase blunted Sph I site.

c. Serotonin $5HT_{2C}$

The cDNA containing a S310K mutation was constructed by replacing the Sty I restriction fragment containing amino acid 310 with synthetic double stranded oligonucleotides that encode

(L258K)		
GPR31 (Q221K)	SEQ.ID.NO.: 195	SEQ.ID.NO.: 196
GPR32 (K255A)	SEQ.ID.NO.: 269	SEQ.ID.NO.: 270
GPR40 (A223K)	SEQ.ID.NO.: 271	SEQ.ID.NO.: 272
GPR41 (A223K)	SEQ.ID.NO.: 273	SEQ.ID.NO.: 274
GPR43 (V221K)	SEQ.ID.NO.: 275	SEQ.ID.NO.: 276
APJ (L247K)	SEQ.ID.NO.: 197	SEQ.ID.NO.: 198
BLR1 (V258K)	SEQ.ID.NO.: 199	SEQ.ID.NO.: 200
CEPR (L258K)	SEQ.ID.NO.: 201	SEQ.ID.NO.: 202
EBI1 (I262K)	SEQ.ID.NO.: 203	SEQ.ID.NO.: 204
EBI2 (L243K)	SEQ.ID.NO.: 205	SEQ.ID.NO.: 206
ETBR-LP2 (N358K)	SEQ.ID.NO.: 207	SEQ.ID.NO.: 208
GHSR (V262K)	SEQ.ID.NO.: 209	SEQ.ID.NO.: 210
GPCR-CNS (N491K)	SEQ.ID.NO.: 211	SEQ.ID.NO.: 212
GPR-NGA (I275K)	SEQ.ID.NO.: 213	SEQ.ID.NO.: 214
H9a (F236K)	SEQ.ID.NO.: 215	SEQ.ID.NO.: 216
H9b (F236K)	SEQ.ID.NO.: 217	SEQ.ID.NO.: 218
HB954 (H265K)	SEQ.ID.NO.: 219	SEQ.ID.NO.: 220
HG38 (V765K)	SEQ.ID.NO.: 277	SEQ.ID.NO.: 278
HM74 (I230K)	SEQ.ID.NO.: 221	SEQ.ID.NO.: 222
MIG (T273K)	SEQ.ID.NO.: 223	SEQ.ID.NO.: 224
OGR1 (Q227K)	SEQ.ID.NO.: 225	SEQ.ID.NO.: 226
Serotonin $5HT_{2A}$ (C322K)	SEQ.ID.NO.: 227	SEQ.ID.NO.: 228
Serotonin $5HT_{2C}$ (S310K)	SEQ.ID.NO.: 229	SEQ.ID.NO.: 230
V28 (I230K)	SEQ.ID.NO.: 231	SEQ.ID.NO.: 232

V28 (I230K)	CAAGAAAGCCAAAGCCAAG AAACTGATCCTTCTG (162)	CTCCTTCGGTCCTCCTATCGT TGTCAGAAGT

The non-endogenous human GPCRs were then sequenced and the derived and verified nucleic acid and amino acid sequences are listed in the accompanying "Sequence Listing" appendix to this patent document, as summarized in Table C below:

Table C

Mutated GPCR	Nucleic Acid Sequence Listing	Amino Acid Sequence Listing
GPR1 (F245K)	SEQ.ID.NO.: 163	SEQ.ID.NO.: 164
GPR4 (K223A)	SEQ.ID.NO.: 165	SEQ.ID.NO.: 166
GPR5 (V224K)	SEQ.ID.NO.: 167	SEQ.ID.NO.: 168
GPR7 (T250K)	SEQ.ID.NO.: 169	SEQ.ID.NO.: 170
GPR8 (T259K)	SEQ.ID.NO.: 171	SEQ.ID.NO.: 172
GPR9 (M254K)	SEQ.ID.NO.: 173	SEQ.ID.NO.: 174
GPR9-6 (L241K)	SEQ.ID.NO.: 175	SEQ.ID.NO.: 176
GPR10 (F276K)	SEQ.ID.NO.: 177	SEQ.ID.NO.: 178
GPR15 (I240K)	SEQ.ID.NO.: 179	SEQ.ID.NO.: 180
GPR17 (V234K)	SEQ.ID.NO.: 181	SEQ.ID.NO.: 182
GPR18 (I231K)	SEQ.ID.NO.: 183	SEQ.ID.NO.: 184
GPR20 (M240K)	SEQ.ID.NO.: 185	SEQ.ID.NO.: 186
GPR21 (A251K)	SEQ.ID.NO.: 187	SEQ.ID.NO.: 188
GPR22 (F312K)	SEQ.ID.NO.: 189	SEQ.ID.NO.: 190
GPR24 (T304K))	SEQ.ID.NO.: 191	SEQ.ID.NO.: 192
GPR30	SEQ.ID.NO.: 193	SEQ.ID.NO.: 194

(A223K)	(266)	
GPR43 (V221K)	GGCGGCGCCGAGCCAAGGGG CTGGCTGTGG (267)	CTCCTTCGGTCCTCCTATCGT TGTCAGAAGT
APJ (L247K)	alternate approach; *see* below	alternate approach; *see* below
BLR1 (V258K)	CAGCGGCAGAAGGCAAAAA GGGTGGCCATC (148)	CTCCTTCGGTCCTCCTATCGT TGTCAGAAGT
CEPR (L258K)	CGGCAGAAGGCGAAGCGCAT GATCCTCGCG (149)	CTCCTTCGGTCCTCCTATCGT TGTCAGAAGT
EBI1 (I262K)	GAGCGCAACAAGGCCAAAA AGGTGATCATC (150)	CTCCTTCGGTCCTCCTATCGT TGTCAGAAGT
EBI2 (L243K)	GGTGTAAACAAAAAGGCTAA AAACACAATTATTCTTATT (151)	CTCCTTCGGTCCTCCTATCGT TGTCAGAAGT
ETBR-LP2 (N358K)	GAGAGCCAGCTCAAGAGCAC CGTGGTG (152)	CTCCTTCGGTCCTCCTATCGT TGTCAGAAGT
GHSR (V262K)	CCACAAGCAAACCAAGAAAA TGCTGGCTGT (153)	CTCCTTCGGTCCTCCTATCGT TGTCAGAAGT
GPCR-CNS (N491K)	CTAGAGAGTCAGATGAAGTG TACAGTAGTGGCAC (155)	CTCCTTCGGTCCTCCTATCGT TGTCAGAAGT
GPR-NGA (I275K)	CGGACAAAAGTGAAAACTAA AAAGATGTTCCTCATT (156)	CTCCTTCGGTCCTCCTATCGT TGTCAGAAGT
H9a and H9b (F236K)	GCTGAGGTTCGCAATAAACT AACCATGTTTGTG (157)	CTCCTTCGGTCCTCCTATCGT TGTCAGAAGT
HB954 (H265K)	GGGAGGCCGAGCTGAAAGCC ACCCTGCTC (158)	CTCCTTCGGTCCTCCTATCGT TGTCAGAAGT
HG38 (V765K)	GGGACTGCTCTATGAAAAAA CACATTGCCCTG (268)	CATCAAGTGTATCATGTGCC AAGTACGCCC (154)
HM74 (I230K)	CAAGATCAAGAGAGCCAAAA CCTTCATCATG (159)	CTCCTTCGGTCCTCCTATCGT TGTCAGAAGT
MIG (T273K)	CCGGAGACAAGTGAAGAAG ATGCTGTTTGTC (160)	CTCCTTCGGTCCTCCTATCGT TGTCAGAAGT
OGR1 (Q227K)	GCAAGGACCAGATCAAGCGG CTGGTGCTCA (161)	CTCCTTCGGTCCTCCTATCGT TGTCAGAAGT
Serotonin $5HT_{2A}$ (C322K)	alternate approach; *see* below	alternate approach; *see* below
Serotonin $5HT_{2C}$ (S310K)	alternate approach; *see* below	alternate approach; *see* below

	(134)	
GPR7 (T250K)	GCCAAGAAGCGGGTGAAGTT CCTGGTGGTGGCA (135)	CTCCTTCGGTCCTCCTATCGT TGTCAGAAGT
GPR8 (T259K)	CAGGCGGAAGGTGAAAGTCC TGGTCCTCGT (136)	CTCCTTCGGTCCTCCTATCGT TGTCAGAAGT
GPR9 (M254K)	CGGCGCCTGCGGGCCAAGCG GCTGGTGGTGGTG (137)	CTCCTTCGGTCCTCCTATCGT TGTCAGAAGT
GPR9-6 (L241K)	CCAAGCACAAAGCCAAGAAA GTGACCATCAC (138)	CTCCTTCGGTCCTCCTATCGT TGTCAGAAGT
GPR10 (F276K)	GCGCCGGCGCACCAAATGCT TGCTGGTGGT (139)	CTCCTTCGGTCCTCCTATCGT TGTCAGAAGT
GPR15 (I240K)	CAAAAAGCTGAAGAAATCTA AGAAGATCATCTTTATTGTCG (140)	CTCCTTCGGTCCTCCTATCGT TGTCAGAAGT
GPR17 (V234K)	CAAGACCAAGGCAAAACGCA TGATCGCCAT (141)	CTCCTTCGGTCCTCCTATCGT TGTCAGAAGT
GPR18 (I231K)	GTCAAGGAGAAGTCCAAAAG GATCATCATC (142)	CTCCTTCGGTCCTCCTATCGT TGTCAGAAGT
GPR20 (M240K)	CGCCGCGTGCGGGCCAAGCA GCTCCTGCTC (143)	CTCCTTCGGTCCTCCTATCGT TGTCAGAAGT
GPR21 (A251K)	CCTGATAAGCGCTATAAAAT GGTCCTGTTTCGA (144)	CTCCTTCGGTCCTCCTATCGT TGTCAGAAGT
GPR22 (F312K)	GAAAGACAAAAGAGAGTCA AGAGGATGTCTTTATTG (145)	CTCCTTCGGTCCTCCTATCGT TGTCAGAAGT
GPR24 (T304K)	CGGAGAAAGAGGGTGAAAC GCACAGCCATCGCC (146)	CTCCTTCGGTCCTCCTATCGT TGTCAGAAGT
GPR30 (L258K)	alternate approach; *see* below	alternate approach; *see* below
GPR31 (Q221K)	AAGCTTCAGCGGGCCAAGGC ACTGGTCACC (147)	CTCCTTCGGTCCTCCTATCGT TGTCAGAAGT
GPR32 (K255A)	CATGCCAACCGGCCCGCGAG GCTGCTGCTGGT (279)	ACCAGCAGCAGCCTCGCGGG CCGGTTGGCATG (280)
GPR40 (A223K)	CGGAAGCTGCGGGCCAAATG GGTGGCCGGC (265)	CTCCTTCGGTCCTCCTATCGT TGTCAGAAGT
GPR41	CAGAGGAGGGTGAAGGGGCT GTTGGCG	CTCCTTCGGTCCTCCTATCGT TGTCAGAAGT

and the 3' primer contained an EcoRI site with the sequence:

5'-TCCGAATTCTCTGTAGACACAAGGCTTTGG-3' (SEQ.ID.NO.: 128)

The 1.1 kb PCR fragment was digested with HindIII and EcoRI and cloned into HindIII-EcoRI site of pCMV expression vector. Nucleic acid (SEQ.ID.NO.: 129) and amino acid (SEQ.ID.NO.: 130) sequences for human V28 were thereafter determined and verified.

Example 2
PREPARATION OF NON-ENDOGENOUS HUMAN GPCRS

1. Site-Directed Mutagenesis

Mutagenesis based upon the Human GPCR Proline Marker approach disclosed herein was performed on the foregoing endogenous human GPCRs using Transformer Site-Directed Mutagenesis Kit (Clontech) according to the manufacturer instructions. For this mutagenesis approach, a Mutation Probe and a Selection Marker Probe (unless otherwise indicated, the probe of SEQ.ID.NO.: 132 was the same throughout) were utilized, and the sequences of these for the specified sequences are listed below in Table B (the parenthetical number is the SEQ. ID.NO.). For convenience, the codon mutation incorporated into the human GPCR is also noted, in standard form:

Table B

Receptor Identifier (Codon Mutation)	Mutation Probe Sequence (5'-3') (SEQ.ID.NO.)	Selection Marker Probe Sequence (5'-3') (SEQ.ID.NO.)
GPR1 (F245K)	GATCTCCAGTAGGCATAAGT GGACAATTCTGG (131)	CTCCTTCGGTCCTCCTATCGT TGTCAGAAG (132)
GPR4 (K223A)	AGAAGGCCAAGATCGCGCGG CTGGCCCTCA (133)	CTCCTTCGGTCCTCCTATCGT TGTCAGAAGT
GPR5 (V224K)	CGGCGCCACCGCACGAAAAA GCTCATCTTC	CTCCTTCGGTCCTCCTATCGT TGTCAGAAGT

sequenced and found to encode two amino acid changes from the published sequences. The first one was a T25N mutation in the N-terminal extracellular domain; the second is an H452Y mutation. Because cDNA clones derived from two independent PCR reactions using Taq polymerase from two different commercial sources (TaqPlus™ from Stratagene and rTth™ Perkin Elmer) contained the same two mutations, these mutations are likely to represent sequence polymorphisms rather than PCR errors. With these exceptions, the nucleic acid (SEQ.ID.NO.: 121) and amino acid (SEQ.ID.NO.: 122) sequences for human $5HT_{2A}$ were thereafter determined and verified.

38. Serotonin $5HT_{2C}$

The cDNA encoding endogenous human $5HT_{2C}$ receptor was obtained from human brain poly-A$^+$ RNA by RT-PCR. The 5' and 3' primers were derived from the 5' and 3' untranslated regions and contained the following sequences:

5'-GACCTCGAGGTTGCTTAAGACTGAAGC-3' (SEQ.ID.NO.: 123)

5'-ATTTCTAGACATATGTAGCTTGTACCG-3' (SEQ.ID.NO.: 124)

Nucleic acid (SEQ.ID.NO.: 125) and amino acid (SEQ.ID.NO.: 126) sequences for human $5HT_{2C}$ were thereafter determined and verified.

39. V28 (GenBank Accession Number: U20350)

The cDNA for human V28 was generated and cloned into pCMV expression vector as follows: PCR was performed using brain cDNA as template and rTth polymerase (Perkin Elmer) with the buffer system provided by the manufacturer, 0.25 μM of each primer, and 0.2 mM of each of the 4 nucleotides. The cycle condition was 30 cycles of: 94°C for 1 min; 65°C for 1min; and 72 °C for 1 min and 20 sec. The 5' PCR primer contained a HindIII site with the sequence:

5'-GGTAAGCTTGGCAGTCCACGCCAGGCCTTC-3' (SEQ.ID.NO.: 127)

The cDNA for human OGR1 was generated and cloned into pCMV expression vector as follows: PCR was performed using genomic DNA as template and rTth polymerase (Perkin Elmer) with the buffer system provided by the manufacturer, 0.25 μM of each primer, and 0.2 mM of each of the 4 nucleotides. The cycle condition was 30 cycles of: 94°C for 1 min; 65°C for 1min; and 72°C for 1 min and 20 sec. The 5' PCR primer was kinased with the sequence:

5'-GGAAGCTTCAGGCCCAAAGATGGGGAACAT-3' (SEQ.ID.NO.: 115):

and the 3' primer contained a BamHI site with the sequence:

5'-GTGGATCCACCCGCGGAGGACCCAGGCTAG -3' (SEQ.ID.NO.: 116).

The 1.1 kb PCR fragment was digested with BamHI and cloned into the EcoRV-BamHI site of pCMV expression vector. Nucleic acid (SEQ.ID.NO.: 117) and amino acid (SEQ.ID.NO.: 118) sequences for human OGR1 were thereafter determined and verified.

37. Serotonin $5HT_{2A}$

The cDNA encoding endogenous human $5HT_{2A}$ receptor was obtained by RT-PCR using human brain poly-A^+ RNA; a 5' primer from the 5' untranslated region with an Xho I restriction site:

5'-GACCTCGAGTCCTTCTACACCTCATC-3' (SEQ.ID.NO: 119)

and a 3' primer from the 3' untranslated region containing an Xba I site:

5'-TGCTCTAGATTCCAGATAGGTGAAAACTTG-3' (SEQ.ID.NO: 120)

PCR was performed using either TaqPlus™ precision polymerase (Stratagene) or rTth™ polymerase (Perkin Elmer) with the buffer system provided by the manufacturers, 0.25 μM of each primer, and 0.2 mM of each of the 4 nucleotides. The cycle condition was 30 cycles of: 94°C for 1 min; 57°C for 1min; and 72°C for 2 min. The 1.5 kb PCR fragment was digested with Xba I and subcloned into Eco RV-Xba I site of pBluescript. The resulting cDNA clones were fully

polymerase (Stratagene) for first round PCR or pfu polymerase (Stratagene) for second round PCR with the buffer system provided by the manufacturer, 0.25 μM of each primer, and 0.2 mM (TaqPlus Precision) or 0.5 mM (pfu) of each of the 4 nucleotides. When pfu was used, 10% DMSO was included in the buffer. The cycle condition was 30 cycles of: 94°C for 1 min; 65°C for 1min; and 72 °C for: (a) 1 min for first round PCR; and (b) 2 min for second round PCR. Because there is an intron in the coding region, two sets of primers were separately used to generate overlapping 5' and 3' fragments. The 5' fragment PCR primers were:

5'-ACCATGGCTTGCAATGGCAGTGCGGCCAGGGGGCACT-3' (external sense) (SEQ.ID.NO.: 109)

and

5'-CGACCAGGACAAACAGCATCTTGGTCACTTGTCTCCGGC-3'(internal antisense) (SEQ.ID.NO.: 110).

The 3' fragment PCR primers were:

5'-GACCAAGATGCTGTTTGTCCTGGTCGTGGTGTTTGGCAT-3' (internal sense) (SEQ.ID.NO.: 111) and

5'-CGGAATTCAGGATGGATCGGTCTCTTGCTGCGCCT-3' (external antisense with an EcoRI site) (SEQ.ID.NO.: 112).

The 5' and 3' fragments were ligated together by using the first round PCR as template and the kinased external sense primer and external antisense primer to perform second round PCR. The 1.2 kb PCR fragment was digested with EcoRI and cloned into the blunt-EcoRI site of pCMV expression vector. Nucleic acid (SEQ.ID.NO.: 113) and amino acid (SEQ.ID.NO.: 114) sequences for human MIG were thereafter determined and verified.

36. OGR1 (GenBank Accession Number: U48405)

and the 3' primer contained an EcoRI site with the sequence:

5'-TTTGAATTCACATATTAATTAGAGACATGG-3' (SEQ.ID.NO.: 262).

The 1.4 kb 3' PCR fragment was digested with EcoRI and subcloned into a blunt-EcoRI site of pCMV vector. The 5' and 3' fragments were then ligated together through a common EcoRV site to generate the full length cDNA clone. Nucleic acid (SEQ.ID.NO.: 263) and amino acid (SEQ.ID.NO.: 264) sequences for human HG38 were thereafter determined and verified.

34. HM74 (GenBank Accession Number: D10923)

The cDNA for human HM74 was generated and cloned into pCMV expression vector as follows: PCR was performed using either genomic DNA or thymus cDNA (pooled) as template and rTth polymerase (Perkin Elmer) with the buffer system provided by the manufacturer, 0.25 μM of each primer, and 0.2 mM of each of the 4 nucleotides. The cycle condition was 30 cycles of: 94°C for 1 min; 65°C for 1min; and 72°C for 1 min and 20 sec. The 5' PCR primer contained an EcoRI site with the sequence:

5'-GGAGAATTCACTAGGCGAGGCGCTCCATC-3' (SEQ.ID.NO.: 105)

and the 3' primer was kinased with the sequence:

5'-GGAGGATCCAGGAAACCTTAGGCCGAGTCC-3' (SEQ.ID.NO.:106).

The 1.3 kb PCR fragment was digested with EcoRI and cloned into EcoRI-SmaI site of pCMV expression vector. Clones sequenced revealed a potential polymorphism involving a N94K change. Aside from this difference, nucleic acid (SEQ.ID.NO.: 107) and amino acid (SEQ.ID.NO.: 108) sequences for human HM74 were thereafter determined and verified.

35. MIG (GenBank Accession Numbers: AFO44600 and AFO44601)

The cDNA for human MIG was generated and cloned into pCMV expression vector as follows: PCR was performed using genomic DNA as template and TaqPlus Precision

vector as follows: PCR was performed using brain cDNA as template and rTth polymerase (Perkin Elmer) with the buffer system provided by the manufacturer, 0.25 μM of each primer, and 0.2 mM of each of the 4 nucleotides. The cycle condition was 30 cycles of 94°C for 1 min, 58°C for 1min and 72 °C for 2 min. The 5' PCR contained a HindIII site with the sequence:

5'-TCCAAGCTTCGCCATGGGACATAACGGGAGCT -3' (SEQ.ID.NO.: 101)

and the 3' primer contained an EcoRI site with the sequence:

5'-CGTGAATTCCAAGAATTTACAATCCTTGCT -3' (SEQ.ID.NO.: 102).

The 1.6 kb PCR fragment was digested with HindIII and EcoRI and cloned into HindIII-EcoRI site of pCMV expression vector. Nucleic acid (SEQ.ID.NO.: 103) and amino acid (SEQ.ID.NO.: 104) sequences for human HB954 were thereafter determined and verified.

33. HG38 (GenBank Accession Number: AF062006)

The cDNA for human HG38 was generated and cloned into pCMV expression vector as follows: PCR was performed using brain cDNA as template and rTth polymerase (Perkin Elmer) with the buffer system provided by the manufacturer, 0.25 μM of each primer, and 0.2 mM of each 4 nucleotides. The cycle condition was 30 cycles of 94°C for 1 min, 56°C for 1min and 72 °C for 1 min and 30 sec. Two PCR reactions were performed to separately obtain the 5' and 3' fragment. For the 5' fragment, the 5' PCR primer contained an HindIII site with the sequence:

5'-CCCAAGCTTCGGGCACCATGGACACCTCCC-3' (SEQ.ID.NO.: 259)

and the 3' primer contained a BamHIsite with the sequence:

5'-ACAGGATCCAAATGCACAGCACTGGTAAGC-3' (SEQ.ID.NO.: 260).

This 5' 1.5 kb PCR fragment was digested with HindIII and BamHI and cloned into an HindIII-BamHI site of pCMV. For the 3' fragment, the 5' PCR primer was kinased with the sequence:

5'-CTATAACTGGGTTACATGGTTTAAC-3' (SEQ.ID.NO. 261)

96) sequences for human GPR-NGA were thereafter determined and verified.

31. H9 (GenBank Accession Number: U52219)

The cDNA for human HB954 was generated and cloned into pCMV expression vector as follows: PCR was performed using pituitary cDNA as template and rTth polymerase (Perkin Elmer) with the buffer system provided by the manufacturer, 0.25 μM of each primer, and 0.2 mM of each 4 nucleotides. The cycle condition was 30 cycles of: 94°C for 1 min, 62°C for 1min and 72 °C for 2 min. The 5' PCR primer contains a HindIII site with the sequence:

5'-GGAAAGCTTAACGATCCCCAGGAGCAACAT-3' (SEQ.ID.NO.: 97)

and the 3' primer contains a BamHI site with the sequence:

5'-CTGGGATCCTACGAGAGCATTTTTCACACAG-3' (SEQ.ID.NO.: 98).

The 1.9 kb PCR fragment was digested with HindIII and BamHI and cloned into HindIII-BamHI site of pCMV expression vector. When compared to the published sequences, a different isoform with 12 bp in frame insertion in the cytoplasmic tail was also identified and designated "H9b." Both isoforms contain two potential polymorphisms involving changes of amino acid P320S and amino acid G448A. Isoform H9a contained another potential polymorphism of amino acid S493N, while isoform H9b contained two additional potential polymorphisms involving changes of amino acid I502T and amino acid A532T (corresponding to amino acid 528 of isoform H9a). Nucleic acid (SEQ.ID.NO.: 99) and amino acid (SEQ.ID.NO.: 100) sequences for human H9 were thereafter determined and verified (in the section below, both isoforms were mutated in accordance with the Human GPCR Proline Marker Algorithm).

32. HB954 (GenBank Accession Number: D38449)

The cDNA for human HB954 was generated and cloned into pCMV expression

polymerase (Perkin Elmer) with the buffer system provided by the manufacturer, 0.25 μM of each primer, and 0.2 mM of each of the 4 nucleotides. The cycle condition was 30 cycles of: 94°C for 1 min; 65°C for 1min; and 72 °C for 2 min. The 5' PCR primer contained a HindIII site with the sequence:

5'-GCAAGCTTGTGCCCTCACCAAGCCATGCGAGCC-3' (SEQ.ID.NO.: 89)

and the 3' primer contained an EcoRI site with the sequence:

5'-CGGAATTCAGCAATGAGTTCCGACAGAAGC-3' (SEQ.ID.NO.: 90).

The 1.9 kb PCR fragment was digested with HindIII and EcoRI and cloned into HindIII-EcoRI site of pCMV expression vector. All nine clones sequenced contained a potential polymorphism involving a S284C change. Aside from this difference, nucleic acid (SEQ.ID.NO.: 91) and amino acid (SEQ.ID.NO.: 92) sequences for human GPCR-CNS were thereafter determined and verified.

30. GPR-NGA (GenBank Accession Number: U55312)

The cDNA for human GPR-NGA was generated and cloned into pCMV expression vector as follows: PCR was performed using genomic DNA as template and rTth polymerase (Perkin Elmer) with the buffer system provided by the manufacturer, 0.25 μM of each primer, and 0.2 mM of each of the 4 nucleotides. The cycle condition was 30 cycles of 94°C for 1 min, 56°C for 1min and 72 °C for 1.5 min. The 5' PCR primer contained an EcoRI site with the sequence:

5'-CAGAATTCAGAGAAAAAAAGTGAATATGGTTTTT-3' (SEQ.ID.NO.: 93)

and the 3' primer contained a BamHI site with the sequence:

5'-TTGGATCCCTGGTGCATAACAATTGAAAGAAT-3' (SEQ.ID.NO.: 94).

The 1.3 kb PCR fragment was digested with EcoRI and BamHI and cloned into EcoRI-BamHI site of pCMV expression vector. Nucleic acid (SEQ.ID.NO.: 95) and amino acid (SEQ.ID.NO.:

site of pCMV expression vector. Nucleic acid (SEQ.ID.NO.: 81) and amino acid (SEQ.ID.NO.: 82) sequences for human ETBR-LP2 were thereafter determined and verified.

28. GHSR (GenBank Accession Number: U60179)

The cDNA for human GHSR was generated and cloned into pCMV expression vector as follows: PCR was performed using hippocampus cDNA as template and TaqPlus Precision polymerase (Stratagene) with the buffer system provided by the manufacturer, 0.25 μM of each primer, and 0.2 mM of each of the 4 nucleotides. The cycle condition was 30 cycles of: 94°C for 1 min; 68°C for 1min; and 72 °C for 1 min and 10 sec. For first round PCR, the 5' PCR primer sequence was:

5'-ATGTGGAACGCGACGCCCAGCG-3' (SEQ.ID.NO.: 83)

and the 3' primer sequence was:

5'-TCATGTATTAATACTAGATTCT-3' (SEQ.ID.NO.: 84).

Two microliters of the first round PCR was used as template for the second round PCR where the 5' primer was kinased with sequence:

5'-TACCATGTGGAACGCGACGCCCAGCGAAGAGCCGGGGT-3'(SEQ.ID.NO.:85)

and the 3' primer contained an EcoRI site with the sequence:

5'-CGGAATTCATGTATTAATACTAGATTCTGTCCAGGCCCG-3'(SEQ.ID.NO.:86).

The 1.1 kb PCR fragment was digested with EcoRI and cloned into blunt-EcoRI site of pCMV expression vector. Nucleic acid (SEQ.ID.NO.: 87) and amino acid (SEQ.ID.NO.: 88) sequences for human GHSR were thereafter determined and verified.

29. GPCR-CNS (GenBank Accession Number: AFO17262)

The cDNA for human GPCR-CNS was generated and cloned into pCMV expression vector as follows: PCR was performed using brain cDNA as template and rTth

vector as follows: PCR was performed using cDNA clone (graciously provided by Kevin Lynch, University of Virginia Health Sciences Center; the vector utilized was not identified by the source) as template and pfu polymerase (Stratagene) with the buffer system provided by the manufacturer supplemented with 10% DMSO, 0.25 μM of each primer, and 0.5 mM of each of the 4 nucleotides. The cycle condition was 30 cycles of: 94°C for 1 min; 60°C for 1min; and 72 °C for 1 min and 20 sec. The 5' PCR primer contained an EcoRI site with the sequence:

5'-CTGGAATTCACCTGGACCACCACCAATGGATA-3' (SEQ.ID.NO.: 75)

and the 3' primer contained a BamHI site with the sequence

5'-CTCGGATCCTGCAAAGTTTGTCATACAG TT-3' (SEQ.ID.NO.: 76).

The 1.2 kb PCR fragment was digested with EcoRI and BamHI and cloned into EcoRI-BamHI site of pCMV expression vector. Nucleic acid (SEQ.ID.NO.: 77) and amino acid (SEQ.ID.NO.: 78) sequences for human EBI2 were thereafter determined and verified.

27. ETBR-LP2 (GenBank Accession Number: D38449)

The cDNA for human ETBR-LP2 was generated and cloned into pCMV expression vector as follows: PCR was performed using brain cDNA as template and rTth polymerase (Perkin Elmer) with the buffer system provided by the manufacturer, 0.25 μM of each primer, and 0.2 mM of each of the 4 nucleotides. The cycle condition was 30 cycles of: 94°C for 1 min; 65°C for 1min; and 72 °C for 1.5 min. The 5' PCR contained an EcoRI site with the sequence:

5'-CTGGAATTCTCCTGCTCATCCAGCCATGCGG -3' (SEQ.ID.NO.: 79)

and the 3' primer contained a BamHI site with the sequence:

5'-CCTGGATCCCCACCCCTACTGGGGCCTCAG -3' (SEQ.ID.NO.: 80).

The 1.5 kb PCR fragment was digested with EcoRI and BamHI and cloned into EcoRI-BamHI

0.2 mM of each of the 4 nucleotides. The cycle condition was 30 cycles of: 94°C for 1 min; 65°C for 1min; and 72°C for 1 min and 20 sec. The 5' PCR primer was kinased with the sequence:

5'-CAAAGCTTGAAAGCTGCACGGTGCAGAGAC-3' (SEQ.ID.NO.:67)

and the 3' primer contained a BamHI site with the sequence:

5'-GCGGATCCCGAGTCACACCCTGGCTGGGCC-3' (SEQ.ID.NO.: 68).

The 1.2 kb PCR fragment was digested with BamHI and cloned into EcoRV-BamHI site of pCMV expression vector. Nucleic acid (SEQ.ID.NO.: 69) and amino acid (SEQ.ID.NO.: 70) sequences for human CEPR were thereafter determined and verified.

25. EBI1 (GenBank Accession Number: L31581)

The cDNA for human EBI1 was generated and cloned into pCMV expression vector as follows: PCR was performed using thymus cDNA as template and rTth polymerase (Perkin Elmer) with the buffer system provided by the manufacturer, 0.25 μM of each primer, and 0.2 mM of each of the 4 nucleotides. The cycle condition was 30 cycles of: 94°C for 1 min; 62°C for 1min; and 72°C for 1 min and 20 sec. The 5' PCR primer contained an EcoRI site with the sequence:

5'-ACAGAATTCCTGTGTGGTTTTACCGCCCAG-3' (SEQ.ID.NO.: 71)

and the 3' primer contained a BamHI site with the sequence:

5'-CTCGGATCCAGGCAGAAGAGTCGCCTATGG-3' (SEQ.ID.NO.: 72).

The 1.2 kb PCR fragment was digested with EcoRI and BamHI and cloned into EcoRI-BamHI site of PCMV expression vector. Nucleic acid (SEQ.ID.NO.: 73) and amino acid (SEQ.ID.NO.: 74) sequences for human EBI1 were thereafter determined and verified.

26. EBI2 (GenBank Accession Number: L08177)

The cDNA for human EBI2 was generated and cloned into pCMV expression

22. APJ (GenBank Accession Number: U03642)

Human APJ cDNA (in pRcCMV vector) was provided by Brian O'Dowd (University of Toronto). The human APJ cDNA was excised from the pRcCMV vector as an EcoRI-XbaI (blunted) fragment and was subcloned into EcoRI-SmaI site of pCMV vector. Nucleic acid (SEQ.ID.NO.: 61) and amino acid (SEQ.ID.NO.: 62) sequences for human APJ were thereafter determined and verified.

23. BLR1 (GenBank Accession Number: X68149)

The cDNA for human BLR1 was generated and cloned into pCMV expression vector as follows: PCR was performed using thymus cDNA as template and rTth polymerase (Perkin Elmer) with the buffer system provided by the manufacturer, 0.25 μM of each primer, and 0.2 mM of each of the 4 nucleotides. The cycle condition was 30 cycles of: 94°C for 1 min; 62°C for 1min; and 72 °C for 1 min and 20 sec. The 5' PCR primer contained an EcoRI site with the sequence:

5'-TGAGAATTCTGGTGACTCACAGCCGGCACAG-3' (SEQ.ID.NO.: 63):

and the 3' primer contained a BamHI site with the sequence:

5'-GCCGGATCCAAGGAAAAGCAGCAATAAAAGG-3' (SEQ.ID.NO.: 64). The 1.2 kb PCR fragment was digested with EcoRI and BamHI and cloned into EcoRI-BamHI site of pCMV expression vector. Nucleic acid (SEQ.ID.NO.: 65) and amino acid (SEQ.ID.NO.: 66) sequences for human BLR1 were thereafter determined and verified.

24. CEPR (GenBank Accession Number: U77827)

The cDNA for human CEPR was generated and cloned into pCMV expression vector as follows: PCR was performed using genomic DNA as template and rTth polymerase (Perkin Elmer) with the buffer system provided by the manufacturer, 0.25 μM of each primer, and

(Perkin Elmer) with the buffer system provided by the manufacturer, 0.25 μM of each primer, and 0.2 mM of each 4 nucleotides. The cycle condition was 30 cycles of 94°C for 1 min, 65°C for 1min and 72 °C for 1 min and 10 sec. The 5' PCR primer contained an HindIII site with the sequence:

5'-CTCAAGCTTACTCTCTCTCACCAGTGGCCAC-3' (SEQ.ID.NO.: 251)

and the 3' primer was kinased with the sequence

5'-CCCTCCTCCCCCGGAGGACCTAGC-3' (SEQ.ID.NO.: 252).

The 1 kb PCR fragment was digested with HindIII and cloned into HindIII-blunt site of pCMV expression vector. Nucleic acid (SEQ.ID.NO.: 253) and amino acid (SEQ.ID.NO.: 254) sequences for human GPR41 were thereafter determined and verified.

21. GPR43 (GenBank Accession Number AF024690)

The cDNA for human GPR43 was generated and cloned into pCMV expression vector as follows: PCR was performed using genomic DNA as template and rTth polymerase (Perkin Elmer) with the buffer system provided by the manufacturer, 0.25 μM of each primer, and 0.2 mM of each 4 nucleotides. The cycle condition was 30 cycles of: 94°C for 1 min; 65°C for 1min; and 72 °C for 1 min and 10 sec. The 5' PCR primer contains an HindIII site with the sequence:

5'-TTTAAGCTTCCCCTCCAGGATGCTGCCGGAC-3' (SEQ.ID.NO.: 255)

and the 3' primer contained an EcoRI site with the sequence:

5'-GGCGAATTCTGAAGGTCCAGGGAAACTGCTA-3' (SEQ.ID.NO. 256).

The 1 kb PCR fragment was digested with HindIII and EcoRI and cloned into HindIII-EcoRI site of pCMV expression vector. Nucleic acid (SEQ.ID.NO.: 257) and amino acid (SEQ.ID.NO.: 258) sequences for human GPR43 were thereafter determined and verified.

sequence:

5'-TAAGAATTCCATAAAAATTATGGAATGG-3' (SEQ.ID.NO.:243)

and the 3' primer contained a BamHI site with the sequence:

5'-CCAGGATCCAGCTGAAGTCTTCCATCATTC-3' (SEQ.ID.NO.: 244).

The 1.1 kb PCR fragment was digested with EcoRI and BamHI and cloned into EcoRI-BamHI site of pCMV expression vector. Nucleic acid (SEQ.ID.NO.: 245) and amino acid (SEQ.ID.NO.: 246) sequences for human GPR32 were thereafter determined and verified.

19. GPR40 (GenBank Accession Number: AF024687)

The cDNA for human GPR40 was generated and cloned into pCMV expression vector as follows: PCR was performed using genomic DNA as template and rTth polymerase (Perkin Elmer) with the buffer system provided by the manufacturer, 0.25 μM of each primer, and 0.2 mM of each 4 nucleotides. The cycle condition was 30 cycles of: 94°C for 1 min, 65°C for 1min and 72 °C for 1 min and 10 sec. The 5' PCR primer contained an EcoRI site with the sequence

5'-GCAGAATTCGGCGGCCCCATGGACCTGCCCCC-3' (SEQ.ID.NO.: 247)

and the 3' primer contained a BamHI site with the sequence

5'-GCTGGATCCCCCGAGCAGTGGCGTTACTTC-3' (SEQ.ID.NO.: 248).

The 1 kb PCR fragment was digested with EcoRI and BamHI and cloned into EcoRI-BamHI site of pCMV expression vector. Nucleic acid (SEQ.ID.NO.: 249) and amino acid (SEQ.ID.NO.: 250) sequences for human GPR40 were thereafter determined and verified.

20. GPR41 (GenBank Accession Number AF024688)

The cDNA for human GPR41 was generated and cloned into pCMV expression vector as follows: PCR was performed using genomic DNA as template and rTth polymerase

using a "TOPO-TA Cloning Kit" (Invitrogen, #K4500-01), following manufacturer instructions. The full-length GPR30 insert was liberated by digestion with BamH1, separated from the vector by agarose gel electrophoresis, and purified using a Sephaglas Bandprep™ Kit (Pharmacia, # 27-9285-01) following manufacturer instructions. The nucleic acid (SEQ.ID.NO.: 55) and amino acid sequence (SEQ.ID.NO.: 56) for human GPR30 were thereafter determined and verified.

17. GPR31 (GenBank Accession Number: U65402)

The cDNA for human GPR31 was generated and cloned into pCMV expression vector as follows: PCR was performed using genomic DNA as template and rTth polymerase (Perkin Elmer) with the buffer system provided by the manufacturer, 0.25 μM of each primer, and 0.2 mM of each of the 4 nucleotides. The cycle condition was 30 cycles of: 94°C for 1 min; 58°C for 1min; and 72 °C for 2 min. The 5' PCR primer contained an EcoRI site with the sequence:

5'-AAGGAATTCACGGCCGGGTGATGCCATTCCC-3' (SEQ.ID.NO.: 57)

and the 3' primer contained a BamHI site with the sequence:

5'-GGTGGATCCATAAACACGGGCGTTGAGGAC -3' (SEQ.ID.NO.: 58).

The 1.0 kb PCR fragment was digested with EcoRI and BamHI and cloned into EcoRI-BamHI site of pCMV expression vector. Nucleic acid (SEQ.ID.NO.: 59) and amino acid (SEQ.ID.NO.: 60) sequences for human GPR31 were thereafter determined and verified.

18. GPR32 (GenBank Accession Number: AF045764)

The cDNA for human GPR32 was generated and cloned into pCMV expression vector as follows: PCR was performed using genomic DNA as template and rTth polymerase (Perkin Elmer) with the buffer system provided by the manufacturer, 0.25 μM of each primer, and 0.2 mM of each 4 nucleotides. The cycle condition was 30 cycles of: 94°C for 1 min; 56°C for 1min; and 72 °C for 1 min and 20 sec. The 5' PCR primer contained an EcoRI site with the

The 1.38 kb PCR fragment was digested with BamHI and cloned into EcoRV-BamHI site of pCMV expression vector. Nucleic acid (SEQ.ID.NO.: 47) and amino acid (SEQ.ID.NO.: 48) sequences for human GPR22 were thereafter determined and verified.

15. GPR24 (GenBank Accession Number: U71092)

The cDNA for human GPR24 was generated and cloned into pCMV expression vector as follows: PCR was performed using genomic DNA as template and rTth polymerase (Perkin Elmer) with the buffer system provided by the manufacturer, 0.25 μM of each primer, and 0.2 mM of each 4 nucleotides. The cycle condition was 30 cycles of: 94°C for 1 min; 56°C for 1min; and 72 °C for 1 min and 20 sec. The 5' PCR primer contains a HindIII site with the sequence:

5'-GTGAAGCTTGCCTCTGGTGCCTGCAGGAGG-3' (SEQ.ID.NO.: 49)

and the 3' primer contains an EcoRI site with the sequence:

5'-GCAGAATTCCCGGTGGCGTGTTGTGGTGCCC-3' (SEQ.ID.NO.: 50).

The 1.3 kb PCR fragment was digested with HindIII and EcoRI and cloned into HindIII-EcoRI site of pCMV expression vector. The nucleic acid (SEQ.ID.NO.: 51) and amino acid sequence (SEQ.ID.NO.: 52) for human GPR24 were thereafter determined and verified.

16. GPR30 (GenBank Accession Number: U63917)

The cDNA for human GPR30 was generated and cloned as follows: the coding sequence of GPR30 (1128bp in length) was amplified from genomic DNA using the primers:

5'-GGCGGATCCATGGATGTGACTTCCCAA-3' (SEQ.ID.NO.: 53) and

5'-GGCGGATCCCTACACGGCACTGCTGAA-3' (SEQ.ID.NO.: 54).

The amplified product was then cloned into a commercially available vector, pCR2.1 (Invitrogen),

PCMV expression vector. Nucleic acid (SEQ.ID.NO.: 39) and amino acid (SEQ.ID.NO.: 40) sequences for human GPR20 were thereafter determined and verified.

13. GPR21 (GenBank Accession Number: U66580)

The cDNA for human GPR21 was generated and cloned into pCMV expression vector as follows: PCR was performed using genomic DNA as template and rTth polymerase (Perkin Elmer) with the buffer system provided by the manufacturer, 0.25 μM of each primer, and 0.2 mM of each of the 4 nucleotides. The cycle condition was 30 cycles of: 94°C for 1 min; 62°C for 1min; and 72°C for 1 min and 20 sec. The 5' PCR primer was kinased with the sequence:

5'-GAGAATTCACTCCTGAGCTCAAGATGAACT-3' (SEQ.ID.NO.: 41)

and the 3' primer contained a BamHI site with the sequence:

5'-CGGGATCCCCGTAACTGAGCCACTTCAGAT-3' (SEQ.ID.NO.: 42).

The 1.1 kb PCR fragment was digested with BamHI and cloned into EcoRV-BamHI site of pCMV expression vector. Nucleic acid (SEQ.ID.NO.: 43) and amino acid (SEQ.ID.NO.: 44) sequences for human GPR21 were thereafter determined and verified.

14. GPR22 (GenBank Accession Number: U66581)

The cDNA for human GPR22 was generated and cloned into pCMV expression vector as follows: PCR was performed using genomic DNA as template and rTth polymerase (Perkin Elmer) with the buffer system provided by the manufacturer, 0.25 μM of each primer, and 0.2 mM of each of the 4 nucleotides. The cycle condition was 30 cycles of: 94°C for 1 min; 50°C for 1min; and 72°C for 1.5 min. The 5' PCR primer was kinased with the sequence:

5'-TCCCCCGGGAAAAAAACCAACTGCTCCAAA-3' (SEQ.ID.NO.: 45)

and the 3' primer contained a BamHI site with the sequence:

5'-TAGGATCCATTTGAATGTGGATTTGGTGAAA-3' (SEQ.ID.NO.: 46).

vector as follows: PCR was performed using genomic DNA as template and rTth polymerase (Perkin Elmer) with the buffer system provided by the manufacturer, 0.25 μM of each primer, and 0.2 mM of each of the 4 nucleotides. The cycle condition was 30 cycles of: 94°C for 1 min; 54°C for 1min; and 72 °C for 1min and 20 sec. The 5' PCR primer was kinased with the sequence:

5'-ATAAGATGATCACCCTGAACAATCAAGAT -3' (SEQ.ID.NO.: 33)

and the 3' primer contained an EcoRI site with the sequence:

5'-TCCGAATTCATAACATTTCACTGTTTATATTGC-3' (SEQ.ID.NO.: 34).

The 1.0 kb PCR fragment was digested with EcoRI and cloned into blunt-EcoRI site of pCMV expression vector. All 8 cDNA clones sequenced contained 4 possible polymorphisms involving changes of amino acid 12 from Thr to Pro, amino acid 86 from Ala to Glu, amino acid 97 from Ile to Leu and amino acid 310 from Leu to Met. Aside from these changes, nucleic acid (SEQ.ID.NO.: 35) and amino acid (SEQ.ID.NO.: 36) sequences for human GPR18 were thereafter determined and verified.

12. GPR20 (GenBank Accession Number: U66579)

The cDNA for human GPR20 was generated and cloned into pCMV expression vector as follows: PCR was performed using genomic DNA as template and rTth polymerase (Perkin Elmer) with the buffer system provided by the manufacturer, 0.25 μM of each primer, and 0.2 mM of each of the 4 nucleotides. The cycle condition was 30 cycles of: 94°C for 1 min; 62°C for 1min; and 72 °C for 1 min and 20 sec. The 5' PCR primer was kinased with the sequence:

5'-CCAAGCTTCCAGGCCTGGGGTGTGCTGG-3' (SEQ.ID.NO.: 37)

and the 3' primer contained a BamHI site with the sequence:

5'-ATGGATCCTGACCTTCGGCCCCTGGCAGA-3' (SEQ.ID.NO.: 38).

The 1.2 kb PCR fragment was digested with BamHI and cloned into EcoRV-BamHI site of

vector. Nucleic acid (SEQ.ID.NO.: 25) and amino acid (SEQ.ID.NO.: 26) sequences for human GPR10 were thereafter determined and verified.

9. GPR15 (GenBank Accession Number: U34806)

The human cDNA sequence for GPR15 was provided in pCDNA3 by Brian O'Dowd (University of Toronto). GPR15 cDNA (1.5kB fragment) was excised from the pCDNA3 vector as a HindIII-Bam fragment and was subcloned into HindIII-Bam site of pCMV vector. Nucleic acid (SEQ.ID.NO.: 27) and amino acid (SEQ.ID.NO.: 28) sequences for human GPR15 were thereafter determined and verified.

10. GPR17 (GenBank Accession Number: Z94154)

The cDNA for human GPR17 was generated and cloned into pCMV expression vector as follows: PCR was performed using genomic DNA as template and rTth polymerase (Perkin Elmer) with the buffer system provided by the manufacturer, 0.25 μM of each primer, and 0.2 mM of each 4 nucleotides. The cycle condition was 30 cycles of: 94°C for 1 min; 56°C for 1min and 72 °C for 1 min and 20 sec. The 5' PCR primer contained an EcoRI site with the sequence:

5'-CTAGAATTCTGACTCCAGCCAAAGCATGAAT-3' (SEQ.ID.NO.: 29)and the 3' primer contained a BamHI site with the sequence:

5'-GCTGGATCCTAAACAGTCTGCGCTCGGCCT-3' (SEQ.ID.NO.: 30).

The 1.1 kb PCR fragment was digested with EcoRI and BamHI and cloned into EcoRI-BamHI site of pCMV expression vector. Nucleic acid (SEQ.ID.NO.: 31) and amino acid (SEQ.ID.NO.: 32) sequences for human GPR17 were thereafter determined and verified.

11. GPR18 (GenBank Accession Number: L42324)

The cDNA for human GPR18 was generated and cloned into pCMV expression

5'-ACGAATTCAGCCATGGTCCTTGAGGTGAGTGACCACCAAGTGCTAAAT-3' (SEQ.ID.NO.: 17)

and the 3' primer contained a BamHI site with the sequence:

5'-GAGGATCCTGGAATGCGGGGAAGTCAG-3' (SEQ.ID.NO.: 18).

The 1.2 kb PCR fragment was digested with EcoRI and cloned into EcoRI-SmaI site of PCMV expression vector. Nucleic acid (SEQ.ID.NO.: 19) and amino acid (SEQ.ID.NO.: 20) sequences for human GPR9 were thereafter determined and verified.

7. GPR9-6 (GenBank Accession Number: U45982)

The cDNA for human GPR9-6 was generated and cloned into pCMV expression vector as follows: PCR was performed using genomic DNA as template and rTth polymerase (Perkin Elmer) with the buffer system provided by the manufacturer, 0.25 µM of each primer, and 0.2 mM of each of the 4 nucleotides. The cycle condition was 30 cycles of: 94°C for 1 min; 62°C for 1min; and 72°C for 1 min and 20 sec. The 5' PCR primer was kinased with the sequence:

5'-TTAAGCTTGACCTAATGCCATCTTGTGTCC-3' (SEQ.ID.NO.: 21)

and the 3' primer contained a BamHI site with the sequence:

5'-TTGGATCCAAAAGAACCATGCACCTCAGAG-3' (SEQ.ID.NO.: 22).

The 1.2 kb PCR fragment was digested with BamHI and cloned into EcoRV-BamHI site of pCMV expression vector. Nucleic acid (SEQ.ID.NO.: 23) and amino acid (SEQ.ID.NO.: 24) sequences for human GPR9-6 were thereafter determined and verified.

8. GPR10 (GenBank Accession Number: U32672)

The human cDNA sequence for GPR10 was provided in pRcCMV by Brian O'Dowd (University of Toronto). GPR10 cDNA (1.3kB fragment) was excised from the pRcCMV vector as an EcoRI-XbaI fragment and was subcloned into EcoRI-XbaI site of pCMV

12) sequences for human GPR7 were thereafter determined and verified.

5. GPR8 (GenBank Accession Number: U22492)

The cDNA for human GPR8 was generated and cloned into pCMV expression vector as follows: PCR was performed using genomic DNA as template and rTth polymerase (Perkin Elmer) with the buffer system provided by the manufacturer, 0.25 µM of each primer, and 0.2 mM of each of the 4 nucleotides. The cycle condition was 30 cycles of: 94°C for 1 min; 62°C for 1min; and 72 °C for 1min and 20 sec. The 5' PCR primer contained an EcoRI site with the sequence:

5'-CGGAATTCGTCAACGGTCCCAGCTACAATG-3' (SEQ.ID.NO.: 13).

and the 3' primer contained a BamHI site with the sequence:

5'-ATGGATCCCAGGCCCTTCAGCACCGCAATAT-3'(SEQ.ID.NO.: 14).

The 1.1 kb PCR fragment was digested with EcoRI and BamHI and cloned into EcoRI-BamHI site of PCMV expression vector. All 4 cDNA clones sequenced contained a possible polymorphism involving a change of amino acid 206 from Arg to Gln. Aside from this difference, nucleic acid (SEQ.ID.NO.: 15) and amino acid (SEQ.ID.NO.: 16) sequences for human GPR8 were thereafter determined and verified.

6. GPR9 (GenBank Accession Number: X95876)

The cDNA for human GPR9 was generated and cloned into pCMV expression vector as follows: PCR was performed using a clone (provided by Brian O'Dowd) as template and pfu polymerase (Stratagene) with the buffer system provided by the manufacturer supplemented with 10% DMSO, 0.25 µM of each primer, and 0.5 mM of each of the 4 nucleotides. The cycle condition was 25 cycles of: 94°C for 1 min; 56°C for 1min; and 72 °C for 2.5 min. The 5' PCR primer contained an EcoRI site with the sequence:

The cDNA for human GPR5 was generated and cloned into pCMV expression vector as follows: PCR was performed using genomic DNA as template and rTth polymerase (Perkin Elmer) with the buffer system provided by the manufacturer, 0.25 μM of each primer, and 0.2 mM of each of the 4 nucleotides. The cycle condition was 30 cycles of: 94°C for 1 min; 64°C for 1min; and 72°C for 1.5 min. The 5' PCR primer contained an EcoRI site with the sequence:

5'-TATGAATTCAGATGCTCTAAACGTCCCTGC-3' (SEQ.ID.NO.: 5)

and the 3' primer contained BamHI site with the sequence:

5'-TCCGGATCCACCTGCACCTGCGCCTGCACC-3' (SEQ.ID.NO.: 6).

The 1.1 kb PCR fragment was digested with EcoRI and BamHI and cloned into EcoRI-BamHI site of PCMV expression vector. Nucleic acid (SEQ.ID.NO.: 7) and amino acid (SEQ.ID.NO.: 8) sequences for human GPR5 were thereafter determined and verified.

4. GPR7 (GenBank Accession Number: U22491)

The cDNA for human GPR7 was generated and cloned into pCMV expression vector as follows: PCR condition- PCR was performed using genomic DNA as template and rTth polymerase (Perkin Elmer) with the buffer system provided by the manufacturer, 0.25 μM of each primer, and 0.2 mM of each of the 4 nucleotides. The cycle condition was 30 cycles of: 94°C for 1 min; 62°C for 1min; and 72°C for 1min and 20 sec. The 5' PCR primer contained a HindIII site with the sequence:

5'-GCAAGCTTGGGGGACGCCAGGTCGCCGGCT-3' (SEQ.ID.NO.: 9)

and the 3' primer contained a BamHI site with the sequence:

5'-GCGGATCCGGACGCTGGGGGAGTCAGGCTGC-3' (SEQ.ID.NO.: 10).

The 1.1 kb PCR fragment was digested with HindIII and BamHI and cloned into HindIII-BamHI site of pCMV expression vector. Nucleic acid (SEQ.ID.NO.: 11) and amino acid (SEQ.ID.NO.:

and amino acid sequences are disclosed herein, those of ordinary skill in the art are credited with the ability to make minor modifications to these sequences while achieving the same or substantially similar results reported below. Particular approaches to sequence mutations are within the purview of the artisan based upon the particular needs of the artisan.

Example 1
Preparation of Endogenous Human GPCRs

A variety of GPCRs were utilized in the Examples to follow. Some endogenous human GPCRs were graciously provided in expression vectors (as acknowledged below) and other endogenous human GPCRs were synthesized *de novo* using publicly-available sequence information.

1. GPR1 (GenBank Accession Number: U13666)

The human cDNA sequence for GPR1 was provided in pRcCMV by Brian O'Dowd (University ofToronto). GPR1 cDNA (1.4kB fragment) was excised from the pRcCMV vector as a NdeI-XbaI fragment and was subcloned into the NdeI-XbaI site of pCMV vector (*see* Figure 3). Nucleic acid (SEQ.ID.NO.: 1) and amino acid (SEQ.ID.NO.: 2) sequences for human GPR1 were thereafter determined and verified.

2. GPR4 (GenBank Accession Numbers: L36148, U35399, U21051)

The human cDNA sequence for GPR4 was provided in pRcCMV by Brian O'Dowd (University ofToronto). GPR1 cDNA (1.4kB fragment) was excised from the pRcCMV vector as an ApaI(blunted)-XbaI fragment and was subcloned (with most of the 5' untranslated region removed) into HindIII(blunted)-XbaI site of pCMV vector. Nucleic acid (SEQ.ID.NO.: 3) and amino acid (SEQ.ID.NO.: 4) sequences for human GPR4 were thereafter determined and verified.

3. GPR5 (GenBank Accession Number: L36149)

known to those in the art and will not be addressed in detail in this patent document.

F. Pharmaceutical Compositions

Candidate compounds selected for further development can be formulated into pharmaceutical compositions using techniques well known to those in the art. Suitable pharmaceutically-acceptable carriers are available to those in the art; for example, *see* Remington's Pharmaceutical Sciences, 16th Edition, 1980, Mack Publishing Co., (Oslo et al., eds.)

G. Other Utility

Although a preferred use of the non-endogenous versions of the disclosed human GPCRs is for the direct identification of candidate compounds as inverse agonists, agonists or partial agonists (preferably for use as pharmaceutical agents), these receptors can also be utilized in research settings. For example, in vitro and in vivo systems incorporating these receptors can be utilized to further elucidate and understand the roles of the receptors in the human condition, both normal and diseased, as well understanding the role of constitutive activation as it applies to understanding the signaling cascade. A value in these non-endogenous receptors is that their utility as a research tool is enhanced in that, because of their unique features, the disclosed receptors can be used to understand the role of a particular receptor in the human body before the endogenous ligand therefor is identified. Other uses of the disclosed receptors will become apparent to those in the art based upon, *inter alia*, a review of this patent document.

EXAMPLES

The following examples are presented for purposes of elucidation, and not limitation, of the present invention. Following the teaching of this patent document that a mutational cassette may be utilized in the IC3 loop of human GPCRs based upon a position relative to a proline residue in TM6 to constitutively activate the receptor, and while specific nucleic acid

b. Go and Gq.

Gq and Go are associated with activation of the enzyme phospholipase C, which in turn hydrolyzes the phospholipid PIP_2, releasing two intracellular messengers: diacycloglycerol (DAG) and inistol 1,4,5-triphoisphate (IP_3). Increased accumulation of IP_3 is associated with activation of Gq- and Go-associated receptors. *See, generally*, "Indirect Mechanisms of Synaptic Transmission," Chpt. 8, From Neuron To Brain (3rd Ed.) Nichols, J.G. et al eds. Sinauer Associates, Inc. (1992). Assays that detect IP_3 accumulation can be utilized to determine if a candidate compound is, *e.g.*, an inverse agonist to a Gq- or Go-associated receptor (*i.e.*, such a compound would decrease the levels of IP_3). Gq-associated receptors can also been examined using an AP1 reporter assay in that Gq-dependent phospholipase C causes activation of genes containing AP1 elements; thus, activated Gq-associated receptors will evidence an increase in the expression of such genes, whereby inverse agonists thereto will evidence a decrease in such expression, and agonists will evidence an increase in such expression. Commercially available assays for such detection are available.

E. Medicinal Chemistry

Generally, but not always, direct identification of candidate compounds is preferably conducted in conjunction with compounds generated via combinatorial chemistry techniques, whereby thousands of compounds are randomly prepared for such analysis. Generally, the results of such screening will be compounds having unique core structures; thereafter, these compounds are preferably subjected to additional chemical modification around a preferred core structure(s) to further enhance the medicinal properties thereof. Such techniques are

constitutively activated GPCRs that couple the Gs protein are associated with increased cellular levels of cAMP. On the other hand, constitutively activated GPCRs that couple the Gi (or Go) protein are associated with decreased cellular levels of cAMP. *See, generally*, "Indirect Mechanisms of Synaptic Transmission," Chpt. 8, From Neuron To Brain (3rd Ed.) Nichols, J.G. et al eds. Sinauer Associates, Inc. (1992). Thus, assays that detect cAMP can be utilized to determine if a candidate compound is, *e.g.*, an inverse agonist to the receptor (*i.e.*, such a compound would decrease the levels of cAMP). A variety of approaches known in the art for measuring cAMP can be utilized; a most preferred approach relies upon the use of anti-cAMP antibodies in an ELISA-based format. Another type of assay that can be utilized is a whole cell second messenger reporter system assay. Promoters on genes drive the expression of the proteins that a particular gene encodes. Cyclic AMP drives gene expression by promoting the binding of a cAMP-responsive DNA binding protein or transcription factor (CREB) which then binds to the promoter at specific sites called cAMP response elements and drives the expression of the gene. Reporter systems can be constructed which have a promoter containing multiple cAMP response elements before the reporter gene, *e.g.*, β-galactosidase or luciferase. Thus, a constitutively activated Gs-linked receptor causes the accumulation of cAMP that then activates the gene and expression of the reporter protein. The reporter protein such as β-galactosidase or luciferase can then be detected using standard biochemical assays (Chen et al. 1995). With respect to GPCRs that link to Gi (or Go), and thus decrease levels of cAMP, an approach to the screening of, *e.g.*, inverse agonists, based upon utilization of receptors that link to Gs (and thus increase levels of cAMP) is disclosed in the Example section with respect to GPR17 and GPR30.

Gq, Gs, Gi, Go) and stimulates release and subsequent binding of GTP to the G protein. The G protein then acts as a GTPase and slowly hydrolyzes the GTP to GDP, whereby the receptor, under normal conditions, becomes deactivated. However, constitutively activated receptors, including the non-endogenous, human constitutively active GPCRs of the present invention, continue to exchange GDP for GTP. A non-hydrolyzable analog of GTP, [^{35}S]GTPγS, can be used to monitor enhanced binding to G proteins present on membranes which express constitutively activated receptors. It is reported that [^{35}S]GTPγS can be used to monitor G protein coupling to membranes in the absence and presence of ligand. An example of this monitoring, among other examples well-known and available to those in the art, was reported by Traynor and Nahorski in 1995. The preferred use of this assay system is for initial screening of candidate compounds because the system is generically applicable to all G protein-coupled receptors regardless of the particular G protein that interacts with the intracellular domain of the receptor.

B 2. Specific GPCR screening assay techniques

C Once candidate compounds are identified using the "generic" G protein-coupled receptor assay (*i.e.*, an assay to select compounds that are agonists, partial agonists, or inverse agonists), further screening to confirm that the compounds have interacted at the receptor site is preferred. For example, a compound identified by the "generic" assay may not bind to the receptor, but may instead merely "uncouple" the G protein from the intracellular domain.

a. Gs and Gi.

Gs stimulates the enzyme adenylyl cyclase. Gi (and Go), on the other hand, inhibit this enzyme. Adenylyl cyclase catalyzes the conversion of ATP to cAMP; thus,

acid for substitution at this 16th position is viable and allows for efficient selection of a non-endogenous amino acid that fits the needs of the investigator. However, as noted, the more preferred non-endogenous amino acids at the 16th position are lysine, hisitidine, arginine and alanine, with lysine being most preferred. Those of ordinary skill in the art are credited with the ability to readily determine proficient methods for changing the sequence of a codon to achieve a desired mutation.

It has also been discovered that occasionally, but not always, the proline residue marker will be preceded in TM6 by W2 (*i.e.*, $W2P^1AA_{15}X$) where W is tryptophan and 2 is any amino acid residue.

Our discovery, amongst other things, negates the need for unpredictable and complicated sequence alignment approaches commonly used by the art. Indeed, the strength of our discovery, while an algorithm in nature, is that it can be applied in a facile manner to human GPCRs, with dexterous simplicity by those in the art, to achieve a unique and highly useful end-product, *i.e.*, a constitutively activated version of a human GPCR. Because many years and significant amounts of money will be required to determine the endogenous ligands for the human GPCRs that the Human Genome project is uncovering, the disclosed invention not only reduces the time necessary to positively exploit this sequence information, but at significant cost-savings. This approach truly validates the importance of the Human Genome Project because it allows for the utilization of genetic information to not only understand the role of the GPCRs in, *e.g.*, diseases, but also provides the opportunity to improve the human condition.

D. Screening of Candidate Compounds

1. Generic GPCR screening assay techniques

When a G protein receptor becomes constitutively active, it couples to a G protein (*e.g.*,

that readily allows for creation of a constitutively active form of the human GPCRs.

As those in the art appreciate, the transmembrane region of a cell is highly hydrophobic; thus, using standard hydrophobicity plotting techniques, those in the art are readily able to determine the TM regions of a GPCR, and specifically TM6 (this same approach is also applicable to determining the EC and IC regions of the GPCR). It has been discovered that within the TM6 region of human GPCRs, a common proline residue (generally near the middle of TM6), acts as a constitutive activation "marker." By counting 15 amino acids from the proline marker, the 16th amino acid (which is located in the IC3 loop), when mutated from its endogenous form to a non-endogenous form, leads to constitutive activation of the receptor. For convenience, we refer to this as the "Human GPCR Proline Marker" Algorithm. Although the non-endogenous amino acid at this position can be any of the amino acids, most preferably, the non-endogenous amino acid is lysine. While not wishing to be bound by any theory, we believe that this position itself is unique and that the mutation at this location impacts the receptor to allow for constitutive activation.

We note that, for example, when the endogenous amino acid at the 16th position is already lysine (as is the case with GPR4 and GPR32), then in order for X to be a non-endogenous amino acid, it must be other than lysine; thus, in those situations where the endogenous GPCR has an endogenous lysine residue at the 16th position, the non-endogenous version of that GPCR preferably incorporates an amino acid other than lysine, preferably alanine, histidine and arginine, at this position. Of further note, it has been determined that GPR4 appears to be linked to Gs and active in its endogenous form (data not shown).

Because there are only 20 naturally occurring amino acids (although the use of non-naturally occurring amino acids is also viable), selection of a particular non-endogenous amino

– nowhere in this art does this problem evidence more annoying exacerbation than in the genetic sequences that encode nucleic acids and proteins. Thus, for consistency and because of the highly unpredictable nature of this art, the following invention is limited, in terms of mammals, to human GPCRs – applicability of this invention to other mammalian species, while a potential possibility, is considered beyond mere rote application.

In general, when attempting to apply common "rules" from one related protein sequence to another or from one species to another, the art has typically resorted to sequence alignment, *i.e.*, sequences are linearized and attempts are then made to find regions of commonality between two or more sequences. While useful, this approach does not always prove to result in meaningful information. In the case of GPCRs, while the general structural motif is identical for all GPCRs, the variations in lengths of the TMs, ECs and ICs make such alignment approaches from one GPCR to another difficult at best. Thus, while it may be desirable to apply a consistent approach to, *e.g.*, constitutive activation from one GPCR to another, because of the great diversity in sequence length, fidelity, etc from one GPCR to the next, a generally applicable, and readily successful mutational alignment approach is in essence not possible. In an analogy, such an approach is akin to having a traveler start a journey at point A by giving the traveler dozens of different maps to point B, without any scale or distance markers on any of the maps, and then asking the traveler to find the shortest and most efficient route to destination B only by using the maps. In such a situation, the task can be readily simplified by having (a) a common "place-marker" on each map, and (b) the ability to measure the distance from the place-marker to destination B – this, then, will allow the traveler to select the most efficient from starting-point A to destination B.

In essence, a feature of the invention is to provide such coordinates within human GPCRs

B. Disease/Disorder Identification and/or Selection

Most preferably, inverse agonists to the non-endogenous, constitutively activated GPCRs can be identified using the materials of this invention. Such inverse agonists are ideal candidates as lead compounds in drug discovery programs for treating diseases related to these receptors. Because of the ability to directly identify inverse agonists, partial agonists or agonists to these receptors, thereby allowing for the development of pharmaceutical compositions, a search, for diseases and disorders associated with these receptors is possible. For example, scanning both diseased and normal tissue samples for the presence of these receptor now becomes more than an academic exercise or one which might be pursued along the path of identifying, in the case of an orphan receptor, an endogenous ligand. Tissue scans can be conducted across a broad range of healthy and diseased tissues. Such tissue scans provide a preferred first step in associating a specific receptor with a disease and/or disorder.

Preferably, the DNA sequence of the endogenous GPCR is used to make a probe for either radiolabeled cDNA or RT-PCR identification of the expression of the GPCR in tissue samples. The presence of a receptor in a diseased tissue, or the presence of the receptor at elevated or decreased concentrations in diseased tissue compared to a normal tissue, can be preferably utilized to identify a correlation with that disease. Receptors can equally well be localized to regions of organs by this technique. Based on the known functions of the specific tissues to which the receptor is localized, the putative functional role of the receptor can be deduced.

C. A "Human GPCR Proline Marker" Algorithm and the Creation of Non-Endogenous, Constitutively-Active Human GPCRs

Among the many challenges facing the biotechnology arts is the unpredictability in gleaning genetic information from one species and correlating that information to another species

intended, nor should be construed, as a limitation on the disclosure or the claims to follow.

A. Introduction

The traditional study of receptors has always proceeded from the a priori assumption (historically based) that the endogenous ligand must first be identified before discovery could proceed to find antagonists and other molecules that could affect the receptor. Even in cases where an antagonist might have been known first, the search immediately extended to looking for the endogenous ligand. This mode of thinking has persisted in receptor research even after the discovery of constitutively activated receptors. What has not been heretofore recognized is that it is the active state of the receptor that is most useful for discovering agonists, partial agonists, and inverse agonists of the receptor. For those diseases which result from an overly active receptor or an under-active receptor, what is desired in a therapeutic drug is a compound which acts to diminish the active state of a receptor or enhance the activity of the receptor, respectively, not necessarily a drug which is an antagonist to the endogenous ligand. This is because a compound that reduces or enhances the activity of the active receptor state need not bind at the same site as the endogenous ligand. Thus, as taught by a method of this invention, any search for therapeutic compounds should start by screening compounds against the ligand-independent active state.

Screening candidate compounds against non-endogenous, constitutively activated GPCRs allows for the direct identification of candidate compounds which act at these cell surface receptors, without requiring any prior knowledge or use of the receptor's endogenous ligand. By determining areas within the body where the endogenous version of such GPCRs are expressed and/or over-expressed, it is possible to determine related disease/disorder states which are associated with the expression and/or over-expression of these receptors; such an approach is disclosed in this patent document.

X_{codon}), the percent sequence homology should be at least 98%.

ORPHAN RECEPTOR shall mean an endogenous receptor for which the endogenous ligand specific for that receptor has not been identified or is not known.

PHARMACEUTICAL COMPOSITION shall mean a composition comprising at least one active ingredient, whereby the composition is amenable to investigation for a specified, efficacious outcome in a mammal (for example, and not limitation, a human). Those of ordinary skill in the art will understand and appreciate the techniques appropriate for determining whether an active ingredient has a desired efficacious outcome based upon the needs of the artisan.

PLASMID shall mean the combination of a Vector and cDNA. Generally, a Plasmid is introduced into a Host Cell for the purpose of replication and/or expression of the cDNA as a protein.

STIMULATE or **STIMULATING**, in relationship to the term "response" shall mean that a response is increased in the presence of a compound as opposed to in the absence of the compound.

TRANSVERSE or **TRANSVERSING**, in reference to either a defined nucleic acid sequence or a defined amino acid sequence, shall mean that the sequence is located within at least two different and defined regions. For example, in an amino acid sequence that is 10 amino acid moieties in length, where 3 of the 10 moieties are in the TM6 region of a GPCR and the remaining 7 moieties are in the IC3 region of the GPCR, the 10 amino acid moiety can be described as transversing the TM6 and IC3 regions of the GPCR.

VECTOR in reference to cDNA shall mean a circular DNA capable of incorporating at least one cDNA and capable of incorporation into a Host Cell.

The order of the following sections is set forth for presentational efficiency and is not

INVERSE AGONISTS shall mean compounds which bind to either the endogenous form of the receptor or to the constitutively activated form of the receptor, and which inhibit the baseline intracellular response initiated by the active form of the receptor below the normal base level of activity which is observed in the absence of agonists or partial agonists, or decrease GTP binding to membranes. Preferably, the baseline intracellular response is inhibited in the presence of the inverse agonist by at least 30%, more preferably by at least 50%, and most preferably by at least 75%, as compared with the baseline response in the absence of the inverse agonist.

KNOWN RECEPTOR shall mean an endogenous receptor for which the endogenous ligand specific for that receptor has been identified.

LIGAND shall mean an endogenous, naturally occurring molecule specific for an endogenous, naturally occurring receptor.

MUTANT or **MUTATION** in reference to an endogenous receptor's nucleic acid and/or amino acid sequence shall mean a specified change or changes to such endogenous sequences such that a mutated form of an endogenous, non-constitutively activated receptor evidences constitutive activation of the receptor. In terms of equivalents to specific sequences, a subsequent mutated form of a human receptor is considered to be equivalent to a first mutation of the human receptor if (a) the level of constitutive activation of the subsequent mutated form of the receptor is substantially the same as that evidenced by the first mutation of the receptor; and (b) the percent sequence (amino acid and/or nucleic acid) homology between the subsequent mutated form of the receptor and the first mutation of the receptor is at least about 80%, more preferably at least about 90% and most preferably at least 95%. Ideally, and owing to the fact that the most preferred cassettes disclosed herein for achieving constitutive activation includes a single amino acid and/or codon change between the endogenous and the non-endogenous forms of the GPCR (i.e. X or

vivo" and "in vitro" systems. For example, and not limitation, in a screening approach, the endogenous or non-endogenous receptor may be in reference to an in vitro screening system whereby the receptor is expressed on the cell-surface of a mammalian cell. As a further example and not limitation, where the genome of a mammal has been manipulated to include a non-endogenous constitutively activated receptor, screening of a candidate compound by means of an in vivo system is viable.

HOST CELL shall mean a cell capable of having a Plasmid and/or Vector incorporated therein. In the case of a prokaryotic Host Cell, a Plasmid is typically replicated as an autonomous molecule as the Host Cell replicates (generally, the Plasmid is thereafter isolated for introduction into a eukaryotic Host Cell); in the case of a eukaryotic Host Cell, a Plasmid is integrated into the cellular DNA of the Host Cell such that when the eukaryotic Host Cell replicates, the Plasmid replicates. Preferably, for the purposes of the invention disclosed herein, the Host Cell is eukaryotic, more preferably, mammalian, and most preferably selected from the group consisting of 293, 293T and COS-7 cells.

INDIRECTLY IDENTIFYING or **INDIRECTLY IDENTIFIED** means the traditional approach to the drug discovery process involving identification of an endogenous ligand specific for an endogenous receptor, screening of candidate compounds against the receptor for determination of those which interfere and/or compete with the ligand-receptor interaction, and assessing the efficacy of the compound for affecting at least one second messenger pathway associated with the activated receptor.

INHIBIT or **INHIBITING**, in relationship to the term "response" shall mean that a response is decreased or prevented in the presence of a compound as opposed to in the absence of the compound.

constitutive activation as compared with the endogenous form of that GPCR, more preferably, about a 25% difference in such comparative response, and most preferably about a 50% difference in such comparative response. When used for the purposes of directly identifying candidate compounds, it is most preferred that the signal difference be at least about 50% such that there is a sufficient difference between the endogenous signal and the non-endogenous signal to differentiate between selected candidate compounds. In most instances, the "difference" will be an increase in signal; however, with respect to Gs-coupled GPCRS, the "difference" measured is preferably a decrease, as will be set forth in greater detail below.

CONTACT or **CONTACTING** shall mean bringing at least two moieties together, whether in an in vitro system or an in vivo system.

DIRECTLY IDENTIFYING or **DIRECTLY IDENTIFIED**, in relationship to the phrase "candidate compound", shall mean the screening of a candidate compound against a constitutively activated G protein-coupled receptor, and assessing the compound efficacy of such compound. This phrase is, under no circumstances, to be interpreted or understood to be encompassed by or to encompass the phrase "indirectly identifying" or "indirectly identified."

ENDOGENOUS shall mean a material that is naturally produced by the genome of the species. ENDOGENOUS in reference to, for example and not limitation, GPCR, shall mean that which is naturally produced by a human, an insect, a plant, a bacterium, or a virus. By contrast, the term **NON-ENDOGENOUS** in this context shall mean that which is not naturally produced by the genome of a species. For example, and not limitation, a receptor which is not constitutively active in its endogenous form, but when mutated by using the cassettes disclosed herein and thereafter becomes constitutively active, is most preferably referred to herein as a "non-endogenous, constitutively activated receptor." Both terms can be utilized to describe both "in

"candidate compound" does not include compounds which were publicly known to be compounds selected from the group consisting of inverse agonist, agonist or antagonist to a receptor, as previously determined by an indirect identification process ("indirectly identified compound"); more preferably, not including an indirectly identified compound which has previously been determined to have therapeutic efficacy in at least one mammal; and, most preferably, not including an indirectly identified compound which has previously been determined to have therapeutic utility in humans.

CODON shall mean a grouping of three nucleotides (or equivalents to nucleotides) which generally comprise a nucleoside (adenosine (A), guanosine (G), cytidine (C), uridine (U) and thymidine (T)) coupled to a phosphate group and which, when translated, encodes an amino acid.

COMPOUND EFFICACY shall mean a measurement of the ability of a compound to inhibit or stimulate receptor functionality, as opposed to receptor binding affinity. A preferred means of detecting compound efficacy is via measurement of, *e.g.*, $[^{35}S]GTP\gamma S$ binding, as further disclosed in the Example section of this patent document.

CONSTITUTIVELY ACTIVATED RECEPTOR shall mean a receptor subject to constitutive receptor activation. In accordance with the invention disclosed herein, a non-endogenous, human constitutively activated G protein-coupled receptor is one that has been mutated to include the amino acid cassette $P^1AA_{15}X$, as set forth in greater detail below.

CONSTITUTIVE RECEPTOR ACTIVATION shall mean stabilization of a receptor in the active state by means other than binding of the receptor with its endogenous ligand or a chemical equivalent thereof. Preferably, a G protein-coupled receptor subjected to constitutive receptor activation in accordance with the invention disclosed herein evidences at least a 10% difference in response (increase or decrease, as the case may be) to the signal measured for

AMINO ACID ABBREVIATIONS used herein are set below:

ALANINE	ALA	A
ARGININE	ARG	R
ASPARAGINE	ASN	N
ASPARTIC ACID	ASP	D
CYSTEINE	CYS	C
GLUTAMIC ACID	GLU	E
GLUTAMINE	GLN	Q
GLYCINE	GLY	G
HISTIDINE	HIS	H
ISOLEUCINE	ILE	I
LEUCINE	LEU	L
LYSINE	LYS	K
METHIONINE	MET	M
PHENYLALANINE	PHE	F
PROLINE	PRO	P
SERINE	SER	S
THREONINE	THR	T
TRYPTOPHAN	TRP	W
TYROSINE	TYR	Y
VALINE	VAL	V

PARTIAL AGONISTS shall mean compounds which activate the intracellular response when they bind to the receptor to a lesser degree/extent than do agonists, or enhance GTP binding to membranes to a lesser degree/extent than do agonists

ANTAGONIST shall mean compounds that competitively bind to the receptor at the same site as the agonists but which do not activate the intracellular response initiated by the active form of the receptor, and can thereby inhibit the intracellular responses by agonists or partial agonists. ANTAGONISTS do not diminish the baseline intracellular response in the absence of an agonist or partial agonist.

CANDIDATE COMPOUND shall mean a molecule (for example, and not limitation, a chemical compound) which is amenable to a screening technique. Preferably, the phrase

Figure 3 is a sequence diagram of the preferred vector pCMV, including restriction enzymen site locations.

Figure 4 is a diagrammatic representation of the signal measured comparing pCMV, non-endogenous, constitutively active GPR30 inhibition of GPR6-mediated activation of CRE-Luc reporter with endogenous GPR30 inhibition of GPR6-mediated activation of CRE-Luc reporter.

Figure 5 is a diagrammatic representation of the signal measured comparing pCMV, non-endogenous, constitutively activated GPR17 inhibition of GPR3-mediated activation of CRE-Luc reporter with endogenous GPR17 inhibition of GPR3-mediated activation of CRE-Luc reporter.

Figure 6 provides diagrammatic results of the signal measured comparing control pCMV, endogenous APJ and non-endogenous APJ.

Figure 7 provides an illustration of IP_3 production from non-endogenous human 5-HT_{2A} receptor as compared to the endogenous version of this receptor.

Figure 8 are dot-blot format results for GPR1 (8A), GPR30 (8B) and APJ (8C).

DETAILED DESCRIPTION

The scientific literature that has evolved around receptors has adopted a number of terms to refer to ligands having various effects on receptors. For clarity and consistency, the following definitions will be used throughout this patent document. To the extent that these definitions conflict with other definitions for these terms, the following definitions shall control:

AGONISTS shall mean compounds that activate the intracellular response when they bind to the receptor, or enhance GTP binding to membranes.

incorporated into mammalian cells and utilized for the screening of candidate compounds, the non-endogenous human GPCR incorporating the mutation need not be purified and isolated *per se* (*i.e.*, these are incorporated within the cellular membrane of a mammalian cell), although such purified and isolated non-endogenous human GPCRs are well within the purview of this disclosure. Gene-targeted and transgenic non-human mammals (preferably rats and mice) incorporating the non-endogenous human GPCRs are also within the purview of this invention; in particular, gene-targeted mammals are most preferred in that these animals will incorporate the non-endogenous versions of the human GPCRs in place of the non-human mammal's endogenous GPCR-encoding region (techniques for generating such non-human mammals to replace the non-human mammal's protein encoding region with a human encoding region are well known; see, for example, U.S. Patent No. 5,777,194.)

It has been discovered that these changes to an endogenous human GPCR render the GPCR constitutively active such that, as will be further disclosed herein, the non-endogenous, constitutively activated version of the human GPCR can be utilized for, *inter alia*, the direct screening of candidate compounds without the need for the endogenous ligand. Thus, methods for using these materials, and products identified by these methods are also within the purview of the following disclosure.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 shows a generalized structure of a G protein-coupled receptor with the numbers assigned to the transmembrane helixes, the intracellular loops, and the extracellular loops.

Figure 2 schematically shows the two states, active and inactive, for a typical G protein coupled receptor and the linkage of the active state to the second messenger transduction pathway.

acid sequences as well. In those cases where the endogenous amino acid at position X is lysine, then in the non-endogenous version of such GPCR, X is an amino acid other than lysine, preferably alanine.

Accordingly, and as a hypothetical example, if the endogenous GPCR has the following endogenous amino acid sequence at the above-noted positions:

P-AACCTTGGRRRDDDE –Q

then any of the following exemplary and hypothetical cassettes would fall within the scope of the disclosure (non-endogenous amino acids are set forth in bold):

P-AACCTTGGRRRDDDE –**K**

P-AACCTT**HI**GRRDDDE –**K**

P-A**DEE**TTGGRRRDDDE –**A**

P-**LLKFMSTWZLVAAPQ** –**K**

A-LLKFMSTWZLVAAPQ –K

It is also possible to add amino acid residues within AA_{15}, but such an approach is not particularly advanced. Indeed, in the most preferred embodiments, the only amino acid that differs in the non-endogenous version of the human GPCR as compared with the endogenous version of that GPCR is the amino acid in position X; mutation of this amino acid itself leads to constitutive activation of the receptor.

Thus, in particularly preferred embodiments, P^1 and P^{codon} are endogenous proline and an endogenous nucleic acid encoding region encoding proline, respectively; and X and X_{codon} are non-endogenous lysine or alanine and a non-endogenous nucleic acid encoding region encoding lysine or alanine, respectively, with lysine being most preferred. Because it is most preferred that the non-endogenous versions of the human GPCRs which incorporate these mutations are

endogenous amino acids, excepting that none of the 15 endogenous codons within the TM6 region of the GPCR encodes a proline amino acid residue; and

(3) X_{codon} is a nucleic acid encoding region residue located within the IC3 region of said GPCR, where X_{codon} encodes a non-endogenous amino acid, preferably selected from the group consisting of lysine, hisitidine and arginine, and most preferably lysine, excepting that when the endogenous encoding region at position X_{codon} encodes the amino acid lysine, then X_{codon} encodes an amino acid other than lysine, preferably alanine.

The terms endogenous and non-endogenous in reference to these sequence cassettes are relative to the endogenous GPCR. For example, once the endogenous proline residue is located within the TM6 region of a particular GPCR, and the 16th amino acid therefrom is identified for mutation to constitutively activate the receptor, it is also possible to mutate the endogenous proline residue (*i.e.*, once the marker is located and the 16th amino acid to be mutated is identified, one may mutate the marker itself), although it is most preferred that the proline residue not be mutated. Similarly, and while it is most preferred that AA_{15} be maintained in their endogenous forms, these amino acids may also be mutated. The only amino acid that must be mutated in the non-endogenous version of the human GPCR is X *i.e.*, the endogenous amino acid that is 16 residues from P^1 cannot be maintained in its endogenous form and must be mutated, as further disclosed herein. Stated again, while it is preferred that in the non-endogenous version of the human GPCR, P^1 and AA_{15} remain in their endogenous forms (*i.e.*, identical to their wild-type forms), once X is identified and mutated, any and/or all of P^1 and AA_{15} can be mutated. This applies to the nucleic

(a) the endogenous GPCR's amino acids (b) non-endogenous amino acid residues, and (c) a combination of the endogenous GPCR's amino acids and non-endogenous amino acids, excepting that none of the 15 endogenous amino acid residues that are positioned within the TM6 region of the GPCR is proline; and

(3) X is a non-endogenous amino acid residue located within the IC3 region of said GPCR, preferably selected from the group consisting of lysine, hisitidine and arginine, and most preferably lysine, excepting that when the endogenous amino acid at position X is lysine, then X is an amino acid other than lysine, preferably alanine;

and/or

(b) P^{codon} (AA-codon)$_{15}$ X_{codon}

wherein:

(1) P^{codon} is a nucleic acid sequence within the TM6 region of the GPCR, where P^{codon} encodes an amino acid selected from the group consisting of (i) the endogenous GPCR's proline residue, and (ii) a non-endogenous amino acid residue other than proline;

(2) (AA-codon)$_{15}$ are 15 codons encoding 15 amino acids selected from the group consisting of (a) the endogenous GPCR's amino acids (b) non-endogenous amino acid residues and (c) a combination of the endogenous GPCR's amino acids and non-

HBA954	1261 Biochim. Biophys. Acta 121 (1995)

HG38	247 Biochem. Biophys. Res. Commun. 266 (1998)
HM74	5 Int. Immunol. 1239 (1993)
OGR1	35 Genomics 397 (1996)
V28	163 Gene 295 (1995)

As will be set forth and disclosed in greater detail below, utilization of a mutational cassette to modify the endogenous sequence of a human GPCR leads to a constitutively activated version of the human GPCR. These non-endogenous, constitutively activated versions of human GPCRs can be utilized, *inter alia*, for the screening of candidate compounds to directly identify compounds of, *e.g.*, therapeutic relevance.

SUMMARY OF THE INVENTION

Disclosed herein is a non-endogenous, human G protein-coupled receptor comprising (a) as a most preferred amino acid sequence region (C-terminus to N-terminus orientation) and/or (b) as a most preferred nucleic acid sequence region (3' to 5' orientation) transversing the transmembrane-6 (TM6) and intracellular loop-3 (IC3) regions of the GPCR:

(a) $P^1\ AA_{15}\ X$

wherein:

(1) P^1 is an amino acid residue located within the TM6 region of the GPCR, where P^1 is selected from the group consisting of (i) the endogenous GPCR's proline residue, and (ii) a non-endogenous amino acid residue other than proline;

(2) AA_{15} are 15 amino acids selected from the group consisting of

otherwise, that the invention disclosed herein is only applicable to orphan GPCRs or the specific GPCRs listed below. Additionally, certain receptors that have been isolated are not the subject of publications per se; for example, reference is made to a G Protein-Coupled Receptor database on the "world-wide web" (neither the named inventors nor the assignee have any affiliation with this site) that lists GPCRs. Other GPCRs are the subject of patent applications owned by the present assignee and these are not listed below (including GPR3, GPR6 and GPR12; *see* U.S. Provisional Number 60/094879):

Table A

Receptor Name	Publication Reference
GPR1	23 Genomics 609 (1994)
GPR4	14 DNA and Cell Biology 25 (1995)
GPR5	14 DNA and Cell Biology 25 (1995)
GPR7	28 Genomics 84 (1995)
GPR8	28 Genomics 84 (1995)
GPR9	184 J. Exp. Med. 963 (1996)
GPR10	29 Genomics 335 (1995)
GPR15	32 Genomics 462 (1996)
GPR17	70 J Neurochem. 1357 (1998)
GPR18	42 Genomics 462 (1997)
GPR20	187 Gene 75 (1997)
GPR21	187 Gene 75 (1997)
GPR22	187 Gene 75 (1997)
GPR24	398 FEBS Lett. 253 (1996)
GPR30	45 Genomics 607 (1997)
GPR31	42 Genomics 519 (1997)
GPR32	50 Genomics 281 (1997)
GPR40	239 Biochem. Biophys. Res. Commun. 543 (1997)
GPR41	239 Biochem. Biophys. Res. Commun. 543 (1997)
GPR43	239 Biochem. Biophys. Res. Commun. 543 (1997)
APJ	136 Gene 355 (1993)
BLR1	22 Eur. J. Immunol. 2759 (1992)
CEPR	231 Biochem. Biophys. Res. Commun. 651 (1997)
EBI1	23 Genomics 643 (1994)
EBI2	67 J. Virol. 2209 (1993)
ETBR-LP2	424 FEBS Lett. 193 (1998)
GPCR-CNS	54 Brain Res. Mol. Brain Res. 152 (1998); 45 Genomics 68 (1997)
GPR-NGA	394 FEBS Lett. 325 (1996)
H9	386 FEBS Lett 219 (1996)

two different conformations: an "inactive" state and an "active" state. As shown schematically in Figure 2, a receptor in an inactive state is unable to link to the intracellular signaling transduction pathway to produce a biological response. Changing the receptor conformation to the active state allows linkage to the transduction pathway (via the G-protein) and produces a biological response.

A receptor may be stabilized in an active state by an endogenous ligand or a compound such as a drug. Recent discoveries, including but not exclusively limited to modifications to the amino acid sequence of the receptor, provide means other than endogenous ligands or drugs to promote and stabilize the receptor in the active state conformation. These means effectively stabilize the receptor in an active state by simulating the effect of an endogenous ligand binding to the receptor. Stabilization by such ligand-independent means is termed "constitutive receptor activation."

As noted above, the use of an orphan receptor for screening purposes has not been possible. This is because the traditional "dogma" regarding screening of compounds mandates that the ligand for the receptor be known. By definition, then, this approach has no applicability with respect to orphan receptors. Thus, by adhering to this dogmatic approach to the discovery of therapeutics, the art, in essence, has taught and has been taught to forsake the use of orphan receptors unless and until the endogenous ligand for the receptor is discovered. Given that there are an estimated 2,000 G protein coupled receptors, the majority of which are orphan receptors, such dogma castigates a creative, unique and distinct approach to the discovery of therapeutics.

Information regarding the nucleic acid and/or amino acid sequences of a variety of GPCRs is summarized below in Table A. Because an important focus of the invention disclosed herein is directed towards orphan GPCRs, many of the below-cited references are related to orphan GPCRs. However, this list is not intended to imply, nor is this list to be construed, legally or

membrane (each span is identified by number, *i.e.*, transmembrane-1 (TM-1), transmebrane-2 (TM-2), etc.). The transmembrane helices are joined by strands of amino acids between transmembrane-2 and transmembrane-3, transmembrane-4 and transmembrane-5, and transmembrane-6 and transmembrane-7 on the exterior, or "extracellular" side, of the cell membrane (these are referred to as "extracellular" regions 1, 2 and 3 (EC-1, EC-2 and EC-3), respectively). The transmembrane helices are also joined by strands of amino acids between transmembrane-1 and transmembrane-2, transmembrane-3 and transmembrane-4, and transmembrane-5 and transmembrane-6 on the interior, or "intracellular" side, of the cell membrane (these are referred to as "intracellular" regions 1, 2 and 3 (IC-1, IC-2 and IC-3), respectively). The "carboxy" ("C") terminus of the receptor lies in the intracellular space within the cell, and the "amino" ("N") terminus of the receptor lies in the extracellular space outside of the cell. The general structure of G protein-coupled receptors is depicted in Figure 1.

Generally, when an endogenous ligand binds with the receptor (often referred to as "activation" of the receptor), there is a change in the conformation of the intracellular region that allows for coupling between the intracellular region and an intracellular "G-protein." Although other G proteins exist, currently, Gq, Gs, Gi, and Go are G proteins that have been identified. Endogenous ligand-activated GPCR coupling with the G-protein begins a signaling cascade process (referred to as "signal transduction"). Under normal conditions, signal transduction ultimately results in cellular activation or cellular inhibition. It is thought that the IC-3 loop as well as the carboxy terminus of the receptor interact with the G protein. A principal focus of this invention is directed to the transmembrane-6 (TM6) region and the intracellular-3 (IC3) region of the GPCR.

Under physiological conditions, GPCRs exist in the cell membrane in equilibrium between

with orphan receptors relative to the discovery of novel therapeutics. This is because the traditional approach to the discovery and development of pharmaceuticals has required access to both the receptor *and* its endogenous ligand. Thus, heretofore, orphan GPCRs have presented the art with a tantalizing and undeveloped resource for the discovery of pharmaceuticals.

Under the traditional approach to the discovery of potential therapeutics, it is generally the case that the receptor is first identified. Before drug discovery efforts can be initiated, elaborate, time consuming and expensive procedures are typically put into place in order to identify, isolate and generate the receptor's endogenous ligand – this process can require from between 3 and ten years per receptor, at a cost of about $5million (U.S.) per receptor. These time and financial resources must be expended before the traditional approach to drug discovery can commence. This is because traditional drug discovery techniques rely upon so-called "competitive binding assays" whereby putative therapeutic agents are "screened" against the receptor in an effort to discover compounds that either block the endogenous ligand from binding to the receptor ("antagonists"), or enhance or mimic the effects of the ligand binding to the receptor ("agonists"). The overall objective is to identify compounds that prevent cellular activation when the ligand binds to the receptor (the antagonists), or that enhance or increase cellular activity that would otherwise occur if the ligand was properly binding with the receptor (the agonists). Because the endogenous ligands for orphan GPCRs are by definition not identified, the ability to discover novel and unique therapeutics to these receptors using traditional drug discovery techniques is not possible. The present invention, as will be set forth in greater detail below, overcomes these and other severe limitations created by such traditional drug discovery techniques.

GPCRs share a common structural motif. All these receptors have seven sequences of between 22 to 24 hydrophobic amino acids that form seven alpha helices, each of which spans the

BACKGROUND OF THE INVENTION

Although a number of receptor classes exist in humans, by far the most abundant and therapeutically relevant is represented by the G protein-coupled receptor (GPCR or GPCRs) class. It is estimated that there are some 100,000 genes within the human genome, and of these, approximately 2% or 2,000 genes, are estimated to code for GPCRs. Of these, there are approximately 100 GPCRs for which the endogenous ligand that binds to the GPCR has been identified. Because of the significant time-lag that exists between the discovery of an endogenous GPCR and its endogenous ligand, it can be presumed that the remaining 1,900 GPCRs will be identified and characterized long before the endogenous ligands for these receptors are identified. Indeed, the rapidity by which the Human Genome Project is sequencing the 100,000 human genes indicates that the remaining human GPCRs will be fully sequenced within the next few years. Nevertheless, and despite the efforts to sequence the human genome, it is still very unclear as to how scientists will be able to rapidly, effectively and efficiently exploit this information to improve and enhance the human condition. The present invention is geared towards this important objective.

Receptors, including GPCRs, for which the endogenous ligand has been identified are referred to as "known" receptors, while receptors for which the endogenous ligand has not been identified are referred to as "orphan" receptors. This distinction is not merely semantic, particularly in the case of GPCRs. GPCRs represent an important area for the development of pharmaceutical products: from approximately 20 of the 100 known GPCRs, 60% of all prescription pharmaceuticals have been developed. Thus, the orphan GPCRs are to the pharmaceutical industry what gold was to California in the late 19th century – an opportunity to drive growth, expansion, enhancement and development. A serious drawback exists, however,

NON-ENDOGENOUS, CONSTITUTIVELY ACTIVATED HUMAN G PROTEIN-COUPLED RECEPTORS

The benefits of commonly owned U.S. Serial Number 09/170,496, filed October 13, 1998, U.S. Serial Number 08/839, 449 filed April 14, 1997 (now abandoned), U.S. Serial Number 09/060,188, filed April 14, 1998; U.S. Provisional Number 60/090,783, filed June 26, 1998; and U.S. Provisional Number 60/095,677, filed on August 7, 1998, are hereby claimed. Each of the foregoing applications are incorporated by reference herein in their entirety.

FIELD OF THE INVENTION

The invention disclosed in this patent document relates to transmembrane receptors, and more particularly to human G protein-coupled receptors (GPCRs) which have been altered such that altered GPCRs are constitutively activated. Most preferably, the altered human GPCRs are used for the screening of therapeutic compounds.